


```

.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: B69257
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-196 <KLE>
A;Cross-references: UNIPROT:O30175; GB:AE001102; GB:AE000782; NID:G2689425; PIDN:AAB9116
C;Superfamily: dTMP kinase

Query Match      6.6%; Score 155; DB 2; Length 196;
Best Local Similarity 28.6%; Pred. NO. 0.00018;
Matches 61; Conservative 37; Mismatches 75; Indels 40; Gaps 10;

Qy 254 VVAIEGLDAGTKTTVTSQVADSLKAVLL-----KAVLLKSP-PSICIGQWRKIFDDEPTIIRAFY 307
Db 2 LIAVEGIDGAGKTTIAVIAELLKEGYKVKVLPKPGDSKFG--KKIKSSEERISPEEL 59

Qy 308 SLGNYIVASET-----AKSPVIVDRYHSTATYATATEVSGGLQHLPPAHPVY 360
Db 60 EL--FLKDRIDARENILPALQSGVAVVMDRYFYSNIAQSGAIDAFI-----IR 108

Qy 361 QWPEDLL-KPDLILLTVSPERLQROGMKTRAEAELEANSVFRQKVMYSQRMEN 419
Db 109 EMNEKIAPKPLTILIDVEPIALERVERKGLSPFEKLD-----YLRKVRKGFLENAD 162

Qy 420 PGCHVVDASPRE-----KVLQTVLSLIONS 445
Db 163 ETTVVVDASKPLEVKEVRKVIESFLNKKNS 195

RESULT 3
F82788
thymidylate kinase XF0580 [imported] - Xylella fastidiosa (strain 945c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
R;anonymus: F82788
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82788
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-217 <SIM>
A;Cross-references: GB:AE003904; GB:AE003849; NID:G9105433; PIDN:AAF83390.1; GSPDB:GN001
A;Experimental source: strain 945c
R;Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Cartaro, D.M.; Carier, H
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, R.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshakao, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
A;Genetics:
A;Gene: XF0580
C;Superfamily: dTMP kinase

Query Match      6.5%; Score 154.5; DB 2; Length 217;
Best Local Similarity 27.7%; Pred. NO. 0.00022;
Matches 59; Conservative 32; Mismatches 71; Indels 51; Gaps 9;

```

```

254 VVAIEGLDAGTKTTVTSQVADSLKAVLL-----KSPSPSCIGQWRKIF-----DD 297
11 LVAIEGIDGAGKTTTAKSLALKURGVLETVVSKPTN--GPGWMLRQSAATGRFSPEE 68
298 EPTIIRAFYSLGNYIVASEIAKESAKSPVIVDRYHSTATYATATEVSGGLQHLPPAHH 357
69 EVDVLLDRDRQHVEDLIVPMIGRAV--VILDRYFSPVWVAY-----QGAAGL----- 113
358 PVYQWPEDELL-----KPDILLTVSPERLQROGMKTRAEAELEANSVFRQK 409
114 -----PVDALLEANAFAPRPDVLILLDVPALGILQRIWERG--STPNHFETTEN---LSR 163
410 VMSYQRMENPGCHVVDASPSREKVLQTVLSLI 442
164 CRDIFLALFLPSKRVKVIDATAETAETVLSAALALV 196

RESULT 4
T41553
thymidylate kinase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T41553
R;Wood, V.; Rajandream, M.A.; Bartell, B.G.; Seeger, K.; Harris, D.
submitted to the EMBL Data Library, June 1998
A;Reference number: 222001
A;Accession: T41553
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-210 <WOO>
A;Cross-references: UNIPROT:P36590; EMBL:AL023794; PIDN:CAA19357.1; GSPDB:GN00068; SPDB:
A;Experimental source: strain 972h-; cosmid c70
C;Genetics:
A;Gene: SPDB:SPCC70.07c
A;Map position: 3
A;Introns: 45/1; 110/3
C;Superfamily: dTMP kinase

Query Match      6.1%; Score 144.5; DB 2; Length 210;
Best Local Similarity 26.1%; Pred. NO. 0.0012;
Matches 58; Conservative 39; Mismatches 76; Indels 49; Gaps 11;

Qy 248 QKGFOVVAIEGLDAGTKTTVTSQVADSL-----KAVLLKSPSPCIGQWRKIFD----- 296
Db 3 KQNGRLIVBELDRSGKSTQCQLLVKLSQHEKAELEKFPDRTTAIGKKIDDYLVKESV 62
297 ---DEPTIIRAFYSLGNYIVASEIAKESAKS-PVIVDRYHSTATYATATEVSGGLQHL 352
63 QLNDQ---VIHLFSANRWETIQIYEQINKGVTCILDY---AFSGIAFSAAGLD-- 113
353 PPAHIFVYQW-----PEDLLKPDILLTVSPERLQROGMKTRAEAELEANSVFRQ 408
114 -----WEMCKSPDRGLPREDLVIFLNVDP--RIATRGQYGEERVEKIEMQ-----E 158
409 KVMSYQRM-----NPGCHVV--DASPSREKVLQTVLSLION 444
159 KVLKNFORLQKFEKREGLFTLDASSSLEVDHVSQIVDLVSN 200

RESULT 5
thymidylate kinase (tmk-1) [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: F90227
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: F90227

```


C;Keywords: ATP; nucleotide binding; nucleotide biosynthesis; p-loop; phosphotransferase F;14-21/Region: nucleotide-binding motif A (p-loop)

```

60  DGSKISYEAEZALLFAADAEHVKIILPALESEKVKICDRYPYSSI----- 105
Db

352 LPPAHHPYQWPEDI-----LKPDILILLTSPERLORLOGRMEKTEEEAE 399
QY

106 -----AYQWARGDLNLWLQVNSFAPRDLAILLDFVKESRIIKLRG-----TITE 153
Db

400 LEANSVPQKVEMSYQRMEN--PGCHVVDASPSREKVIQTVLISLQN 444
QY

154 FDKIVELORKVRNYLKLAEAFPEMKRIVNAJSSIEDHSIDIVALVKH 200
Db

```

```

RESULT 9
A82128
thymidylate kinase VC2016 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: A82128
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Charadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bacs, S.; Qin, H.; Dragoi, I.; Sellers,
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: A82128
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <HEI>
A:Cross-references: UNIPROT:Q9KQI2; GB:AE004276; GB:AE003852; NID:g9656555; PIDN:AAF9516
A:Experimental source: serogroup O1; strain N16961, biotype El Tor
C:Genetics:
A:Gene: VC2016
A:Map position: 1
C:Superfamily: dtmP kinase

```

	Query Match	5.3%; Score 124.5; DB 2; Length 212;
	Best Local Similarity	24.1%; Pred. No. 0.04;
	Matches	53; Conservative 36; Mismatches 88; Indels 43; Gaps 7;
Qy	255 VAIEGIDATGKTTTQSV-----ADSLKAVLLKSPSCIGOW	291
	: :: :	: :: :
Db	6 IVIEGLEGAKSTAIQVVETLQQNGIDHITRPEPGTILAEKLRAIVKEHPG-----	60
Qy	292 RKTFD-DEPTIRAFYSLGNVIVASIAKESAPVIVDRYWHSTTYAIAITEVSGGLQ	350
	: :: :	: :: :
Db	61 EELQDITELLIVAAARVOLVENVIKPALARGEW--VVGDRHDMSOAYQ-----GGGRQ	112
Qy	351 HLPFAHPHVQPWPEDLLKPDILLTLVTSPPERLQLQGMEKTRFEAELEANSVPKQV	410
	: :: :	: :: :
Db	113 IASTWSLKQTALGDGFKPDLTLYDDPKLGIERARGRGELDRIEKWDIS----FFERA	168
Qy	411 EMSYQRMENPGCHV--DASPSEKVLTQVLISIONSFSE	448
Db	169 RERYELELANDSDSVMIMDAAQSBQVTADIRRALQDWLSO	208

RESULT 10
S28955:
dtmp kinase (EC 2.7.4.9) - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 09-Jul-2004
C:accession: S28955
R:Baigam, L.T.; Yeh, Y.I.S.; Jong, A.Y.
Biochim. Biophys. Acta 1132, 222-234, 1992
A:title: Functional and structural conservation of Schizosaccharomyces pombe
A:reference number: S28955; PMID:93003330; PMID:1327149
A:accession: S28955
A:molecule type: mRNA
A:Residues: 1-210 >ABA>
A:cross-references: UNIPROT:P36590; EMBL:X65868; NID:54973; PIDN:CAA46598.1
C:Superfamily: dtmp kinase

Query Match	5.2%;	Score 123.5;	DB 2;	Length 210;
Best local Similarity	24.3%;	Pred. No. 0.047;		
Matches	57;	Conservative	36;	Mismatches 67; Indels 75; Gaps 12
Qy	248	QKGFQVVALEGD	ATGKTTVTQSVADSL-----KAVLLKSPSPSIGOWRKIFDD-----	297
Db	3	KQNGRLIVIEGL	DRSKSTQCQLLVKLIILNMKRLKLFKFPDRTAIGKKI--DDYLTS 61	
Qy	298	-----	EPILIRAFYSLGNVIYVASEIAKESAKSPVIVDRYWHSTANYA 340	
Db	62	VQLNDQVIHLLF	SANRWEPSIYRA-----NQRCNCILDRY-----AFSG 102	
Qy	341	IATEVSGGLQHL	PPAHHPVQW---PEDLLKPLDILLTVSPERLQRLQGRGMKTR 396	
Db	103	IATSAAKGLD-----	WEWKSPDRGLTRPDLVIFLNVDP--RIANATRGQGEERYE 151	
Qy	397	EABLEANSVPQK	VMYSQYRME-----NPGCHV--DASP-SREKVLQTVLSLQN 444	
Db	152	KIEMQ-----	EKVLKLQRLQREFREGLEFFITLDASSVALEDVDSQIVLNSN 200	

RESULT 11

D90518
thymidylate kinase (dtmp kinase) [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: D90518
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pul
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: D90518
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <KUR>
A:Cross-references: UNIPROT:Q98RF7; GR:AL445566; PID:gl4089465; PIDN:CAC13225.1; GSPDB:
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPU_0520
C:Superfamily: dtmp kinase

Query Match	5.2%	Score 123;	DB 2;	Length 234;
Best Local Similarity	25.5%;	Pred. No. 0.059;		
Matches	56;	Conservative 43;	Mismatches 75;	Indels 46; Gaps 14
Qy	255	VATEGIDATCKTTVTQSVADSLKAVLLKS---	PPSCIGOW---	PK-----IPDDE 298
Db	16	ITEGIDASCKT-----SLLAKLAVQKNLQKCTETWPGGRKSPSEIQIRHLLNKE 71		
Qy	299	PTI--IRRAF-YSLG-----NYIVASIAKESAKSPVIVRYWHSTATY-AIAETVSGL 349		
Db	72	SNLSPIAEAPLYSSARRIHDKVILNLAKNV---VFCRFVDSSPAYQAFGRDL--GF 126		
Qy	350	QHLPPAHPVYQWPEDLLKEDLILLNVSPPEERLQRLQGRGMEXTREAEAEANSVFRQK 409		
Db	127	EKTLN-----ELATDKYDITFIKLISYSESMERKAR--QDEDRLEKEENS-FVQK 179		
Qy	410	VENSYORMEN-----PGCHVVDASPSREKVLQTVLSLIQN 444		
Db	180	VIKGYDFLASYPFQRRIFVIDASKNOEEIFESVLKLD 219		

```

RESULT 12
S66058
thymidylate kinase tmk - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S66059; D69724
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.

```


DNA Res. 1, 1-14, 1994
A;Title: Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome
A;Reference number: S65967; MUID:96051385; PMID:7584024
A;Accession: S66058
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-212 <OGA>
A;Cross-references: UNIPROT:P37537; EMBL:D26185; NID:9467326; PIDN:BAA05264.1; PID:g4674
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
R;Kunst, F.; Ogasawara, S.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Brucher, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emeksson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
A.; Koether, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A.; Authors: Yoshikawa, H.F.; Zunstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: D69724
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-212 <KUN>
A;Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11804.1; PID:g2632295
A;Experimental source: strain 168
C;Genetics: tmk
C;Superfamily: dTWP kinase
C;Keywords: nucleotide binding; P-loop
F;10-17/Region: nucleotide-binding motif A (P-loop)

Query Match 5.1%; Score 120; DB 2; Length 212;
Best Local Similarity 25.0%; Pred. No. 0.087;
Matches 53; Conservative 36; Mismatches 97; Indels 26; Gaps 8;
Qy 255 VAIGLDATGKTTTQSVADSLKA-----VLKSPSCIG---OWRKIFDDEPTII-----302
Db 6 ITFEGEGAGKTVLQETKNLTAEGLQVMAATREPGGDIAEQIREVILNENNILMDPKT 65
Qy 303 RRAEYSLGNYIVASIEIAKESAKSPVIV--DRYWHSTATYATATEVSGLOGLHPAHPV 360
Db 66 EALLYAAARROHLVEKVKPALEQGFVLCDFIDSPAYQGVARGLGIDEVL-----SIN 120
Qy 361 QWPEDLLKPDILLITVSPERLQGRGMEKTRAEALEANSV-FRQKVMYSYORMEN 419
Db 121 EFAIGDMPHVTYVFSIDPEGLKRIYANG---SREKNRLDLEKIDFHTKVQEGYQELMK 177
Qy 420 ---PGCHVVDASPREKVLQTVLSLIQNSFSF 448
Db 178 RFPERFHSVDAGQSKDLVVQDVLKVIDEALKK 209

RESULT 13
KIVZSW
dTMP kinase (EC 2.7.4.9) - vaccinia virus
N;Alternate names: A48R protein; SalFlIR protein; thymidylate kinase
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 31-Mar-1991 #sequence revision 31-Mar-1991 #text_change 09-Jul-2004
C;Accession: E42522; JQ1786; S06181; T37434
R;Goebel, S.J.; Johnson, G.P.; Perkins, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 517-563, 1990
A;Title: Appendix to "The complete DNA sequence of vaccinia virus".
A;Reference number: A42501
A;Accession: E42522
A;Molecule type: DNA
A;Residues: 1-204 <GOE>

A;Cross-references: UNIPROT:P13410; GB:M35027; NID:g335317; PIDN:AAA48180.1; PID:g335526
A;Experimental source: strain Copenhagen
R;Smith, G.L.; Chan, Y.S.; Howard, S.T.
J. Gen. Virol. 72, 1349-1376, 1991
A;Title: Nucleotide sequence of 42kbp of vaccinia virus strain WR from near the right i
A;Reference number: JQ1767; MUID:91259063; PMID:2045793
A;Accession: JQ1786
A;Molecule type: DNA
A;Residues: 1-204 <SMI>
A;Cross-references: DDBJ:D11079; NID:g222717; PIDN:BAA01822.1; PID:g222737
A;Experimental source: strain strain WR
R;Goebel, S.J.; Johnson, G.P.; Perkins, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 247-266, 1990
A;Title: The complete DNA sequence of vaccinia virus.
A;Reference number: A42531; MUID:91021027; PMID:2219722
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
R;Smith, G.L.; de Carlos, A.; Chan, Y.S.
Nucleic Acids Res. 17, 7581-7590, 1989
A;Title: Vaccinia virus encodes a thymidylate kinase gene: sequence and transcriptional
A;Reference number: S06181; MUID:90016845; PMID:2552411
A;Accession: S06181
A;Molecule type: DNA
A;Residues: 1-204 <SM2>
A;Cross-references: EMBL:X16259; NID:g62231; PIDN:CAA34345.1; PID:g62232
R;Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dörner, F.
submitted to the EMBL Data Library, March 1997
A;Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) stra
A;Reference number: Z20877
A;Accession: T37434
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-204 <ANT>
A;Cross-references: EMBL:U94848; PIDN:AAB96539.1
A;Experimental source: strain Ankara
C;Genetics: tmk
A;Note: MVA16.R
C;Superfamily: dTWP kinase
C;Keywords: ATP; nucleotide binding; nucleotide biosynthesis; P-loop; phosphotransferas
F;11-18/Region: nucleotide-binding motif A (P-loop)
F;17/Binding site: ATP (Lys) #status predicted

Query Match 5.1%; Score 119.5; DB 1; Length 204;
Best Local Similarity 28.0%; Pred. No. 0.09;
Matches 51; Conservative 28; Mismatches 66; Indels 37; Gaps 8;
Qy 254 VVAIEGLDATGKTTTQSVADSLKAVLLK--SPSCIGQWRKIFDDEPTIIRAFYSLGN 311
Db 6 LIVFEGLDKSGKTTQCNWIMESIPANTIKYLPQRSIVTGKMDIDYLT--RKTYN--D 61
Qy 312 YIV-----ASEIAKESAKS-PVIVDRYWHSTATYATATEVSGLOGLHPAHPV 359
Db 62 HIVNLLFCANRWFASFQIEQEQITLIVDRYAFSGVAYAAAGASMTLSK-----113
Qy 360 YQWPEDLLKPDILLITVSPERLQGRGMEKTRAEALEANSVFRQKVMYSYORMEN 419
Db 114 -SYESGLPKDLVIFLE--SGSKEINRNNGE-----EIVEDVTFOOKVLEFKKME 162
Qy 420 PG 421
Db 163 EG 164

RESULT 14
F64336
dTMP kinase (EC 2.7.4.9) - Methanococcus jannaschii
N;Alternate names: thymidylate kinase
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: F64336
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2004, 20:21:58 ; Search time 157 Seconds
(without alignments)
1645.498 Million cell updates/sec

Title: US-10-681-223-2

Perfect score: 2362

Sequence: 1 MAFARLLRGLSGPLIGRR.....SREKVLQTVLSLQNSFSEP 449

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1684	71.3	417	2 Q6PFG7	Q6pfg7 mus musculus
2	1684	71.3	417	2 AAH57565	AAH57565 mus muscu
3	1552	65.7	395	2 Q9DC34	Q9dc34 mus musculus
4	1466	62.1	303	2 Q6ZRU2	Q6zru2 homo sapien
5	1466	62.1	303	2 BAC87217	BAC87217 homo sapi
6	677.5	28.7	431	2 Q62316	Q62316 mus musculu
7	301	12.7	274	2 Q60970	Q60970 leishmania
8	291	12.3	58	2 Q96AL8	Q96al8 homo sapien
9	169.5	7.2	227	1 KTHY_XANAC	Q9pfg7 xanthomonas
10	166.5	7.0	227	1 KTHY_XANCP	Q9p3y6 xanthomonas
11	155	6.6	196	1 KTHY_ARCFU	Q30175 archaeoglob
12	154.5	6.5	208	1 KTHY_XYLFA	Q9pfs7 xylella fas
13	148.5	6.3	208	1 KTHY_XYLFT	Q97b89 xylella fas
14	145	6.1	202	2 Q6L016	Q6l016 picophilus
15	144.5	6.1	210	1 KTHY_SCHPO	P36590 schizosacch
16	143	6.1	189	1 KTHI_SULSO	Q9uxg7 sulfolobus
17	142	6.0	204	2 Q72NL9	Q72nl9 leptospira
18	142	6.0	204	2 AAS71368	Aas71368 leptospir
19	140	5.9	209	1 KTHY_PASMU	Q9ckes pasteurella
20	137.5	5.8	263	2 Q9LEB4	Q9leb4 arabidopsis
21	135.5	5.7	210	1 KTHY_PYRFU	Q9u071 pyrococcus
22	134	5.7	191	2 Q8SS22	Q8ss22 encephalito
23	131.5	5.6	205	2 Q6N1S4	Q6nis4 corynebacte
24	131.5	5.6	205	2 CAE49210	CAe49210 corynebac
25	130.5	5.5	199	1 KTHY_TROW8	Q93hb8 tropheryma
26	130.5	5.5	199	1 KTHY_TROWT	Q93fm3 tropheryma
27	130.5	5.5	206	1 KTHY_METWA	Q9pxv5 methanosarc
28	129.5	5.5	211	2 Q6F0E1	Q6foe1 mesoplasma
29	129	5.5	211	2 Q9AOD5	Q9aod5 lactococcus
30	129	5.5	213	2 Q6MU15	Q6mu15 mycoplasma
31	129	5.5	213	2 CAE76699	CAe76699 mycoplasm

32 127.5 5.4 190 1 KTHI_SULTO Q975e6 sulfolobus
33 127.5 5.4 205 1 KTHY_PYRAB Q9vie9 pyrococcus
34 127 5.4 204 2 Q6CC22 Q6cc22 yarrowia li
35 127 5.4 1082 1 IPO4_MOUSE Q8vi75 mus musculus
36 126 5.3 210 1 KTHY_HAEIN P44719 haemophilus
37 126 5.3 218 2 Q6FSB3 Q6fsb3 candida gla
38 125.5 5.3 205 1 KTHY_PYRHO O59366 pyrococcus
39 124.5 5.3 212 1 KTHY_VIBCH Q9xq12 vibrio chol
40 124 5.2 208 2 Q6CW62 Q6cw62 kluyveromyc
41 124 5.2 227 2 Q80HT9 Q80ht9 vaccinia vi
42 123.5 5.2 213 1 KTHY_ECOL6 Q8fin9 escherichia
43 123 5.2 227 2 Q8ODS7 Q8ods7 cowpox viru
44 123 5.2 234 1 KTHY_MYCFU Q98rf7 mycoplasma
45 123 5.2 1017 1 UB35_HUMAN Q9p2h5 homo sapien

ALIGNMENTS

RESULT 1
Q6PFG7 PRELIMINARY; PRT; 417 AA.
ID AC Q6PFG7
AD Q6PFG7; 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Tyki protein.
GN Name=Tyki;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=C57BL/6; TISSUE=Mouse;
RC MEDLINE=22398257; Pubmed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=C57BL/6; TISSUE=Mouse;
RC Strausberg R.;
RA Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC057565; AAH57565.1; -.
DR InterPro; IPR000062; Thymidylate kin.
DR Pfam; PF02223; Thymidylate kin; 1.
SQ SEQUENCE 417 AA; 46646 MW; 978DE9D0AA83F12A CRC64;

Query Match 71.3%; Score 1684; DB 2; Length 417;

Best Local Similarity 77.1%; Pred. No. 3.5e-115;

Matches 323; Conservative 38; Mismatches 54; Indels 4; Gaps 2;

Qy 27 MAPPCRVLLEPDCITLAHFALGADAPGDADAPPRLAALLGPPERSYSLCVPTPDAGCG 86

Db 1 MARPRRTVELPDCSLTHFVLG-DG---TTHRDLARLAELGPPGRSVALCVLPAPGEGCG 56

Qy	27	MAPP	CRFVLELPDCTTLAHFALGADAPGDADAPDRLAALLGPPERSYSICVPTPDAGCG	86
Db	1	MAR	PRFTVELPDCSLTHFVLG-DA---TDHDAARLAELLGPPGRSYALCVPLAPGEGCG	56
Qy	87	ARV	AARLHORLLHQLRRGPPRCQQLRLCYCPGQAGGAGGQOQGLLRDPLDDPDRQAL	146
Db	57	PRV	QAARVHRLQLLRGPPRCQQLSLGVPGDQAGEAOGFLLRDPCDHPDTRDL	116
Qy	147	LEL	GACQEARPHLGEFEADPRGQLWQRLWEVDGRRLQVGCACVVPVDEPHLHPVVD	206
Db	117	LQL	LSGCQEARPQLAEFQADSGLLWQRLWEQDGRQVQVDCACVPAQEPHLPLLED	176
Qy	207	LP	SVVFPDREAAARVLEECTSFIPPEARAVLDLVDQCPKQIQKGFQVVAIEGLDATGKT	266
Db	177	LL	NSAVFQDRDAARVLEECTSFIPPEARAVLDLVDQCPKEVKQFQVIAIEGLDATGKT	236
Qy	267	TVT	QSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFYSLGNIVASIAKSAKSP	326
Db	237	TLT	QSVESLKAVALQSPPCISQWRKIFDDEPTIIRAFYSLGNIVASIAKSTNFP	296
Qy	327	VIV	DYWHSTATYAIATEVSGGLQHLPPAHPVYQWMPEDLLKPDLLLTVPSEERLQRL	386
Db	297	VIV	DYWHSTATYAIATEVSGGQYLPFAHPVYQWPGDLLKFDLVLLLTVNSEERVRL	356
Qy	387	QGR	MEKTRBAEAELEANSVFRQKVMYSORMENPGCHVVDASPSREKVLQTVLSLIONS	445
Db	357	QGR	GQKTKBAEAELEANNVFRQKVMYSORMENPSCHLVVDASPSRETVLQKVLLELIQSS	415
RESULT 3				
Q9DC34				
ID	Q9DC34	PRELIMINARY;	PRT;	395 AA.
AC	Q9DC34;			
DT	01-JUN-2001	(TREMELrel. 17, Created)		
DT	01-JUN-2001	(TREMELrel. 17, Last sequence update)		
DT	01-OCT-2003	(TREMELrel. 25, Last annotation update)		
DE	Mus musculus adult male lung cDNA, RIKEN full-length enriched library,			
DE	clone:120004E04 product:thymidylate kinase family LPS-inducible			
DE	member, full insert sequence.			
GN	Name=Tyki.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Lung;			
RX	MEDLINE=99279253; PubMed=10349636;			
RA	Carninci P., Hayashizaki Y.;			
RT	"High-efficiency full-length cDNA cloning.";			
RL	Meth. Enzymol. 303:19-44(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Lung;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	RIKEN FANTOM Consortium;			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Lung;			
RA	The FANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RT	60,770 full-length cDNAs.";			
RL	Nature 420:563-573(2002).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Lung;			
RX	MEDLINE=20499374; PubMed=11042159;			
RA	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,			
RA	Komno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;			

87	ARV	AARLHORLLHQLRRGPPRCQQLRLCYCPGQAGGAGGQOQGLLRDPLDDPDRQAL	146
57	PRV	QAARVHRLQLLRGPPRCQQLSLGVPGDQAGEAOGFLLRDPCDHPDTRDL	116
147	LEL	GACQEARPHLGEFEADPRGQLWQRLWEVDGRRLQVGCACVVPVPPPLHPVVPD	206
117	LQL	LSGCQEARPQLAEFQADSGLLWQRLWEQDGRQVQVDCACVPAQEPHLPLLPD	176
207	LP	SVVFPDREAAARVLEECTSFIPPEARAVLDLVDQCPKQIQKGFQVVAIEGLDATGKT	266
177	LL	NSAVFQDRDAARVLEECTSFIPPEARAVLDLVDQCPKEVKQFQVIAIEGLDATGKT	236
267	TVT	QSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFYSLGNIVASIAKSAKSP	326
237	TLT	QSVESLKAVALQSPPCISQWRKIFDDEPTIIRAFYSLGNIVASIAKSTNFP	296
327	VIV	DYWHSTATYAIATEVSGGLQHLPPAHPVYQWMPEDLLKPDLLLTVPSEERLQRL	386
297	VIV	DYWHSTATYAIATEVSGGLQYLPFAHPVYQWPGDLLKFDLVLLLTVNSEERVRL	356
387	QGR	MEKTRBAEAELEANSVFRQKVMYSORMENPGCHVVDASPSREKVLQTVLSLIONS	445
357	QGR	GQKTKBAEAELEANNVFRQKVMYSORMENPSCHLVVDASPSRETVLQKVLLELIQSS	415
RESULT 2			
AAH57565			
ID	AAH57565	PRELIMINARY;	PRT; 417 AA.
AC	AAH57565;		
DT	02-MAR-2004	(TREMELrel. 27, Created)	
DT	02-MAR-2004	(TREMELrel. 27, Last sequence update)	
DT	02-MAR-2004	(TREMELrel. 27, Last annotation update)	
DE	Tyki protein.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=Mouse;		
RX	MEDLINE=22388257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,		
RA	Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzyszinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=Mouse;		
RA	Strausberg R.;		
RL	Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC057565; AAH57565.1; -		
SQ	SEQUENCE 417 AA; 46646 MW; 978DE9D0A83F12A CRC64;		
Query Match 71.3%; Score 1684; DB 2; Length 417;			
Best Local Similarity 77.1%; Pred. No. 3.5e-115;			
Matches 323; Conservative 38; Mismatches 54; Indels 4; Gaps 2;			

RT RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP "Normalization and subtraction of cap-trapper-selected cDNAs to
RC prepare full-length cDNA libraries for rapid discovery of new genes."
RX [5]
RY STRAIN=C57BL/6J; TISSUE=Lung;
RZ MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Katsunai T., Tashiro H., Itoh M.,
RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RA "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer."
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanaoka T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami T., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toyota T., Yamamura M., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK004595; BAB23396.1; -
DR HGSP; P00572; 1TMK.
DR MGD; MGI-99830; Tyki.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004798; F:thymidylate kinase activity; IEA.
DR GO; GO:0006233; P:dTTP biosynthesis; IEA.
DR GO; GO:0006235; P:dTTP biosynthesis; IEA.
DR InterPro; IPR000062; Thymidylate kin.
DR Pfam; PF02223; Thymidylate_kin; I.
KW Kinase.
SQ SEQUENCE 395 AA; 44085 MW; 62F338B649983DF6 CRC64;
Query Match 65.7%; Score 1552; DB 2; Length 395;
Best Local Similarity 77.8%; Pred. No. 1.6e-105;
Matches 294; Conservative 36; Mismatches 48; Indels 0; Gaps 0;
Qy 68 PPSRSLCVPTPDAGCGARVRAARLHQLRRLGPFQRCOLLRLCYCPGQAGGA 127
Db 16 PQGAATRLCVLAPGEGCGPRVQARVHRLLQQLRRLGRLQRCQLSLGPGDQAGEA 75
Qy 128 QQGLRLDPLDDPTROALLLELGGACQAPRPHLGEFEADPRGOLWRLWEVQGRRLQV 187
Db 76 QHGLRLDPCDHPDTRRLDQLLGLSCQEAARPAQAEFQADSGQLLWRLWEVQGRRLQV 135
Qy 188 GCAQVVPVPEPLHPVPPDPSVVPFDPREARAVLEECTSFIPPEARAVLDVDCPKQI 247
Db 136 DCACVLPQAEPLHPPLLDLNSAVFQDRAARAVLEECTSFIPPEARAVLDVDCPKVEV 195
Qy 248 QKGQFQVVAIEGLDAGTKTTVQSVADSLKAVLLKSPSCIGQWRKIFDDEPTTIIRAFY 307
Db 196 QKGQFQVVAIEGLDAGTKTTVQSVADSLKAVLLKSPSCIGQWRKIFDDEPTTIIRAFY 255
Qy 308 SLGNVIVASEIAKESAPVTVDRVWHTATVAIATEVSGGLQHPAHHPVQWPEDL 367
Db 256 SLGNVIVASEIAKESAPVTVDRVWHTATVAIATEVSGGLQHPAHHPVQWPEDL 315
Qy 368 KPDILLTVPSEERLQLRGMEKTRAEAELEANSVFRQKVENSMENPGCHVDA 427
Db 316 KPDILLTVPSEERLQLRGMEKTRAEAELEANSVFRQKVENSMENPGCHVDA 375

Qy 428 SPSREKVLQTVLSLIONS 445
Db 376 SPSRETVLQKVLLEIQSS 393
RESULT 4
Q6ZRU2 PRELIMINARY; PRT; 303 AA.
ID Q6ZRU2
AC Q6ZRU2
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein FLJ46098.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Kanehori K., Ishibashi T., Chiba Y., Fujimori K., Hiraoka S.,
RA Tanai H., Watanabe S., Ishida S., Ono Y., Hotuta T., Watanabe M.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK127983; BAC87217.1; -
DR GO; GO:0016301; F:kinase activity; IEA.
KW Kinase.
SQ SEQUENCE 303 AA; 32645 MW; 4C7D0715B7B970C6 CRC64;
Query Match 62.1%; Score 1466; DB 2; Length 303;
Best Local Similarity 99.6%; Pred. No. 2.3e-99;
Matches 278; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MAFARRLLRGLSLGRRGVCAGAVPPCRFVLELPCDTLAHFALGADAPGDADPDP 60
Db 1 MAFARRLLRGLSLGRRGVCAGAVPPRRFVLELPCDTLAHFALGADAPGDADPDP 60
Qy 61 RLAAALLGPPERSYSLCVPTPDAGCGARVRAARLHQLRRLGPFQRCOLLRLCYCP 120
Db 61 RLAAALLGPPERSYSLCVPTPDAGCGARVRAARLHQLRRLGPFQRCOLLRLCYCP 120
Qy 121 GQAGGAGQGFLLRDLDDPTROALLLELGGACQAPRPHLGEFEADPRGOLWRLWEVQ 180
Db 121 GQAGGAGQGFLLRDLDDPTROALLLELGGACQAPRPHLGEFEADPRGOLWRLWEVQ 180
Qy 181 DGRLOVGCQAVPVPPFPLHPVPPDPSVVPFDPREARAVLEECTSFIPPEARAVLDV 240
Db 181 DGRLOVGCQAVPVPPFPLHPVPPDPSVVPFDPREARAVLEECTSFIPPEARAVLDV 240
Qy 241 DQCFKQIKGKFOVVAIEGLDAGTKTTVQSVADSLKAV 279
Db 241 DQCFKQIKGKFOVVAIEGLDAGTKTTVQSVADSLKAV 279
RESULT 5
BAC87217 PRELIMINARY; PRT; 303 AA.
ID BAC87217
AC BAC87217
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DE CDNA FUJ46098 fls. clone TESTI2021654, weakly similar to Mus musculus
DE thymidylate kinase family LPS-inducible member (Tyki).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.


```

Q96AL8
ID Q96AL8 PRELIMINARY; PRT; 58 AA.
AC Q96AL8;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
KW EMBL; BC016969; AAH16969.1;
FT NON_TER
SQ SEQUENCE 58 AA; 653 MW; C12029FED37F4BAC CRC64;

Query Match 12.3%; Score 291; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 392 EKTREAELEANSVFRQKVMSTORMENPGCHVVDASPSREKVLQTVLSLQNSFSSEP 449
DB 1 EKTREAELEANSVFRQKVMSTORMENPGCHVVDASPSREKVLQTVLSLQNSFSSEP 58

RESULT 9
KTHY_XANAC
ID KTHY_XANAC STANDARD; PRT; 227 AA.
AC Q8PF67;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Thymidylate kinase (EC 2.7.4.9) (dTMP Kinase).
GN Name=tmk; OrderedLocusNames=XAC4014;
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).

-!- FUNCTION: Phosphorylation of dTMP to form dTDP in both de novo and
CC salvage pathways of dTTP synthesis (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + dTMP = ADP + dTDP.
CC -!- SIMILARITY: Belongs to the thymidylate kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE012049; AA038850.1;
CC HSSP; O05891; 1N5L.
CC HMAP; MF_00165; 1.
CC InterPro; IPR000062; Thymidylate kin.
CC Pfam; PF02223; Thymidylate kin; 1.
CC TIGRFAMs; TIGR00041; DTMP Kinase; 1.
CC PROSITE; PS01331; THYMIDYLATE KINASE; FALSE NEG.
KW ATP-binding; Complete proteome; Kinase; Nucleotide biosynthesis;
KW Transferase. 16 23 ATP (Potential).
FT NP_BIND 227 AA; 24030 MW; 7C78722F1BCFA211 CRC64;
SQ SEQUENCE 227 AA; 24030 MW; 7C78722F1BCFA211 CRC64;

Query Match 7.2%; Score 169.5; DB 1; Length 227;
Best Local Similarity 27.4%; Pred. No. 0.00023;
Matches 59; Conservative 37; Mismatches 68; Indels 49; Gaps 8;

QY 254 VVAIEGLDAGKTVTVQSVADSLKA-----VLLKSPSCIGQW-----RKLFDD 297
DB 11 LIAIEGIDGAGKTVTLARLAAATLDAAGARVYLSKEPTN--GPMGTQLRQSAATGRLSAE 68

QY 298 EPTIIRAFYSLGNYIVASFIKESAKSPVIVDRYWHSTATYATVATVSVGLQHLRPAHH 357
DB 69 EASLLRDRHEHVDTLIAPALAGDI---VILDRYFMSVAYQ-----GAGLPL--- 114

QY 358 PVYQWPEDLL-----KPDLIILLTVSPERLQRLQGRMEKTRBEEAELEANSVFRQKV 410
DB 115 -----LDELLERNAFAPRPDVVLLLDLPPTGLARIRARGDAPNHEFTQDNL-----ERC 164

QY 411 EMSYORMENPGCHVVDASPSREKVLQTVLSLI 442
DB 165 RTTFAALELPGKHVDASDADSVLRQAHAI 196

RESULT 10
KTHY_XANCP
ID KTHY_XANCP STANDARD; PRT; 227 AA.
AC Q8P3Y6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Thymidylate kinase (EC 2.7.4.9) (dTMP Kinase).
GN Name=tmk; OrderedLocusNames=XCC3931;
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).

```


CC -!- FUNCTION: Phosphorylation of dTMP to form dTDP in both de novo and
 CC salvage pathways of dTTP synthesis (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + dTMP = ADP + dTDP.
 CC -!- SIMILARITY: Belongs to the thymidylate kinase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE012513; AAM43153.1; -.
 CC HSSP; P37345; STMP.
 CC InterPro; IPR000062; Thymidylate_kin.
 CC Pfam; PF02223; Thymidylate_kin; 1.
 CC TIGRFAMs; TIGR00041; DTMP_Kinase; 1.
 CC PROSITE; PS01331; THYMIDYLATE_KINASE; FALSE NEG.
 CC ATP-binding; Complete proteome; Kinase; Nucleotide biosynthesis;
 KW Transferase.
 FT NP_BIND 16 23 ATP (Potential).
 SQ SEQUENCE 227 AA; 24000 MW; BDB0D3C93E66753C CRC64;
 Query Match 7.0%; Score 166.5; DB 1; Length 227;
 Best Local Similarity 26.8%; Pred. No. 0.00038;
 Matches 57; Conservative 35; Mismatches 70; Indels 51; Gaps 8;
 QY 254 VVAIEGLDAGTKTTVTQSVADSL-----KAVLLKSPSCIGQW-----RKTFDD 297
 DB 11 LIAIEGDAGKTTTARSATLLAQAGARVVLSKEPTN--GPWGTQLRQSAATGRLSAQD 68
 QY 298 EPTTIIRAFYSLGNYIVASEITAKSAPSVIVDRYHSTATYATATEVSGGLQHLPPAHH 357
 DB 69 EVDLLDRRHVEALIPALARGEI---VILDRYFPMWAY---QGAAGL----- 113
 QY 358 PVYQWPEDDL-----KPDILLITVSPERLQGRGMEKTRAEAELEANSVPFQK 409
 DB 114 -----PDLALUANDFAPRPDLLILLDUPPTGLARIRAGDAPNHFTQNL-----ER 163
 QY 410 VEMSYQRMENPGCHVDASPREKVLQTVLSLI 442
 DB 164 CRAIFPALQLPGKHVIDASADSVLRQAHVV 196
 RESULT 11
 KTHY_XYLCFU
 ID KTHY_ARCFU STANDARD; PRT; 196 AA.
 AC O30175;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
 GN Name=tmk; OrderedLocustNames=AF0061;
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=984975; DOI=10.1038/37052;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.J.,
 RA Peterson S.N., Reich C.I., McNeil L.K., Badger J.H., Glodek A.,
 RA Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,
 RA Utterback T.R., Cotton M.D., Spriggs T., Artiach P., Kaine B.P.,
 RA Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,
 RA Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,

Woese C.R., Venter J.C.;
 "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon Archaeoglobus fulgidus";
 Nature 390:364-370(1997).
 CC -!- CATALYTIC ACTIVITY: ATP + dTMP = ADP + dTDP.
 CC -!- SIMILARITY: Belongs to the thymidylate kinase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001102; AAB91163.1; -.
 CC PIR; B69257; E69257.
 CC HSSP; P37345; STMP.
 CC TIGR; AF0061; -.
 CC HAMAP; MF 00165; -; 1.
 CC InterPro; IPR000062; Thymidylate_kin.
 CC Pfam; PF02223; Thymidylate_kin; 1.
 CC TIGRFAMs; TIGR00041; DTMP_Kinase; 1.
 CC PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
 CC ATP-binding; Complete proteome; Kinase; Nucleotide biosynthesis;
 KW Transferase.
 FT NP_BIND 7 14 ATP (Potential).
 SQ SEQUENCE 196 AA; 22349 MW; F9CF20D740A5A367 CRC64;
 Query Match 6.6%; Score 155; DB 1; Length 196;
 Best Local Similarity 28.6%; Pred. No. 0.0022;
 Matches 61; Conservative 37; Mismatches 75; Indels 40; Gaps 10;
 QY 254 VVAIEGLDAGTKTTVTQSVADSL-----KAVLLKSP-PSGIGQWRKIFDDEPTTIIRAFY 307
 DB 2 LIAVEGDAGKTTTAAATYIAELLKEKGKGVKVLKPGDSKFG--KKIKSSEERLSPBEL 59
 QY 308 SLGNYIVASEI-AKES-----AKSPVIVDRYHSTATYATATEVSGGLQHLPPAHHPVY 360
 DB 60 EL--FLKDRIDARENILPALQSGYAVVMDRYEYNATYQSGARGIDARL-----IR 108
 QY 361 QWPEDLL-KPDILLITVSPERLQGRGMEKTRAEAELEANSVPFQKVEYMSYQRMEN 419
 DB 109 EMNEKIAPKPDLLTLLDVEPIALERVKRGKLSPFKEKLD-----YLRKVRKCFLENAD 162
 QY 420 PGCHVDASPRE-----KVLQTVLSLIONS 445
 DB 163 ETTVVVDASKPLEEVKKEVRKVFESFLNLKKN 195
 RESULT 12
 KTHY_XYLFA
 ID KTHY_XYLFA STANDARD; PRT; 208 AA.
 AC Q9PFS7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
 GN Name=tmk; OrderedLocustNames=XF0580;
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; PubMed=10910347; DOI=10.1038/35018003;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvares M.H., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barrios R., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,

RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A., Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R., RA Garner M.B., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A., RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P., RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C., RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A., RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L., RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y., RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B., RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S., RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A., RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B., RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M., RA da Silva V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E., RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr., RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A., RA de Souza A.P., Terenzi M.P., Truffi D., Tsai S.M., Tsuchako M.H., RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L., RA Zago M.A., Zatz M., Zeldanis J., Setubal J.C.;
 RL Nature 406:151-159(2000).
 CC -!- FUNCTION: Phosphorylation of dTMP to form dTDP in both de novo and
 CC salvage pathways of dTTP synthesis (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + dTMP = ADP + dTDP.
 CC -!- SIMILARITY: Belongs to the thymidylate kinase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE003904; AAF83390.1; ALT_INIT.
 DR HAMAP; MF 00165; -. 1.
 DR InterPro; IPR000062; Thymidylate kin.
 DR Pfam; PF02223; Thymidylate kin; 1.
 DR TIGRFAMs; TIGR00041; dTMP kinase; 1.
 DR PROSITE; PS01331; THYMIDYLATE_KINASE; FALSE NEG.
 KW ATP-binding; Complete proteome; Kinase; Nucleotide biosynthesis;
 KW Transferase.
 FT NP_BIND 7 14 ATP (Potential).
 SQ SEQUENCE 208 AA; 22436 MW; 5563ADA4F9D541AE CRC64;
 Query Match 6.5%; Score 154.5; DB 1; Length 208;
 Best Local Similarity 27.7%; Pred. No. 0.0026;
 Matches 59; Conservative 32; Mismatches 71; Indels 51; Gaps 9;
 QY 254 VVAIEGLDAGTKTTVTSQVADSLKAVLL-----KSPSPCIGQWRKIF-----DD 297
 Db 2 LVAIEGIDGAGKTTLARSALKRGVGLTWSKEPTN--GPWGMLLRQSAATGRFSPEE 59
 QY 298 EPTIIRAFYSLGNYIVASEIAKESAKSPVIVDHYWHESTATYATATEVSGGLQHLPPAHH 357
 Db 60 EVDVLLDRDRQHVEDLIVPMIGRAV---VILDRYFPPSWAY-----QGAAGL----- 104
 QY 358 EPTIIRAFYSLGNYIVASEIAKESAKSPVIVDHYWHESTATYATATEVSGGLQHLPPAHH 409
 Db 105 -----PVDALLEANAFAPRDPVLLLDVPPVIGLQRIWERG--STPNHFETTEN-----LSR 154
 QY 410 VMSYORMENPGCHVVDASPSREKVLQTVLSLI 442
 Db 155 CRDIFLALPLSKRVIDATANAETVLSAALV 187
 RESULT 13
 ID_KTHY_XYLFT STANDARD; PRT; 208 AA.
 AC Q87B89;
 DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
 GN Name:tmk; OrderedLocusNames=PD1569;
 OS Xylella fastidiosa (strain Temecual / ATCC 700964).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=183190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22421331; PubMed=12533478;
 RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
 RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
 RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
 RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorriy H., Tsai S.M.,
 RA Carver H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
 RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
 RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
 RA Baia G.S., Blanco S.R., Brito M.S., Cannavan P.S., Celestino A.V.,
 RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
 RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.F., Sena J.A.D.,
 RA de Souza A.L., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
 RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
 RA Kitajima J.P.;
 RT "Comparative analyses of the complete genome sequences of Pierce's
 RT disease and citrus variegated chlorosis strains of Xylella
 RT fastidiosa.";
 RT J. Bacteriol. 185:1018-1026(2003).
 RL J. Bacteriol. 185:1018-1026(2003).
 CC -!- FUNCTION: Phosphorylation of dTMP to form dTDP in both de novo and
 CC salvage pathways of dTTP synthesis (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + dTMP = ADP + dTDP.
 CC -!- SIMILARITY: Belongs to the thymidylate kinase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE012559; AAO29411.1; -.
 DR HSSP; P37345; STMP.
 DR HAMAP; MF 00165; -. 1.
 DR InterPro; IPR000062; Thymidylate kin.
 DR Pfam; PF02223; Thymidylate kin; 1.
 DR PROSITE; PS01331; THYMIDYLATE_KINASE; FALSE NEG.
 KW ATP-binding; Complete proteome; Kinase; Nucleotide biosynthesis;
 KW Transferase.
 FT NP_BIND 7 14 ATP (Potential).
 SQ SEQUENCE 208 AA; 22455 MW; 254E0C9268E33633 CRC64;
 Query Match 6.3%; Score 148.5; DB 1; Length 208;
 Best Local Similarity 27.2%; Pred. No. 0.0072;
 Matches 58; Conservative 31; Mismatches 73; Indels 51; Gaps 9;
 QY 254 VVAIEGLDAGTKTTVTSQVADSLKAVLL-----KSPSPCIGQWRKIF-----DD 297
 Db 2 LVAIEGIDGAGKTTLARSALKRGVGLTWSKEPTN--GPWGTLRLQSAVTGRFSPEE 59
 QY 298 EPTIIRAFYSLGNYIVASEIAKESAKSPVIVDHYWHESTATYATATEVSGGLQHLPPAHH 357
 Db 60 EVDVLLDRDRQHVEDLIVPMIGRAV---VILDRYFPPSWAY-----QGAAGL----- 104
 QY 358 EPTIIRAFYSLGNYIVASEIAKESAKSPVIVDHYWHESTATYATATEVSGGLQHLPPAHH 409
 Db 105 -----PVDALLEANAFAPRDPVLLLDVPPVIGLQRIWERG--STPNHFETTEN-----LSR 154
 QY 410 VMSYORMENPGCHVVDASPSREKVLQTVLSLI 442
 Db 155 CRDIFLALPLSKRVIDATANAETVFSALGLV 187

RESULT 14

Q6L016 PRELIMINARY; PRT; 202 AA.
 AC Q6L016;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Putative thymidylate kinase (EC 2.7.4.9).
 GN Ordered locus names=PROL101;
 OS Schizosaccharomyces pombe.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Picophilaceae; Picophilus.
 OX NCBI_TaxID=92076;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 9790 / ATCC 700027;
 RX PubMed=15184674;
 RA Fuetterer O., Angelov A., Liesegang H., Gottschalk G., Schleper C.,
 RA Schepers B., Dock C., Antranikian G., Liebl W.;
 RT "Genome sequence of *Picrophilus torridus* and its implications for life
 RT around pH 0.";
 RL Proc. Natl. Acad. Sci. U.S.A. 101:9091-9096 (2004).
 DR EMBL; AE017261; AA43686.1; -;
 DR GO; GO:0016301; F:kinase activity; IEA.
 DR GO; GO:0004798; F:thymidylate kinase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR000062; Thymidylate_kin.
 DR Pfam; PF02223; Thymidylate_kin; 1.
 DR TIGRFAMS; TIGR00041; dTMP_kinase; 1.
 DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
 KW Complete proteome; Kinase; Transferase.
 SQ SEQUENCE 202 AA; 23548 MW; ECD5A5FC43B67EB6 CRC64;

Query Match 6.1%; Score 145; DB 2; Length 202;
 Best Local Similarity 30.0%; Pred. No. 0.012;
 Matches 45; Conservative 22; Mismatches 71; Indels 12; Gaps 4;

QY 255 VAIEGLDATGKTTVTSQVADSLKAVLLKSPS--CIGQWRKIFDEPTIIRAFVSLGNY 312
 DB 9 IALIGDGSKTTUAGDIASVTGYLTREPTDRFCYDIADYDESSINFFLTLDRY 68
 QY 313 IVASEIAKSAKSPVIVDRVHSTATYATATEVSGGLQHLPPAHPVYQWPE-----LLK 368
 DB 69 MHQKEI-KNHLINGVSDRVVFSIAQV-----GSGMEKRFKNWDETISWMLDVSRLIM 122
 QY 369 PDLILLITVSPERLQRLQGRGMKTRBAA 398
 DB 123 PDLIIYLKIDPGIALKRLNLKNEKNTDA 152

RESULT 15

KTHY_SCHPO STANDARD; PRT; 210 AA.
 ID KTHY_SCHPO
 AC P36590; 074528;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
 GN Name=tmpl; Synonyms=tmpl; ORFNames=SPCC70.07c;
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93003330; PubMed=13271149;
 RA Abaigar L.T., Yeh Y.I., Jong A.Y.;
 RT "Functional and structural conservation of *Schizosaccharomyces pombe*
 RT dTMP kinase gene.";
 RL Biochim. Biophys. Acta 1132:222-224 (1992).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne R., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odel C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Stevens K., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grynoprez B.,
 RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Spakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of *Schizosaccharomyces pombe*.";
 RL Nature 415:871-880 (2002).
 CC -!- FUNCTION: Catalyzes the conversion of dTMP to dTDP.
 CC -!- CATALYTIC ACTIVITY: ATP + dTMP = ADP + dTDP.
 CC -!- PATHWAY: Biosynthesis of dTMP from dTDP.
 CC -!- SIMILARITY: Belongs to the thymidylate kinase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; X65868; CAA46698.1; -;
 DR EMBL; AL023794; CAA19357.1; -;
 DR PIR; S28955; S28955.
 DR PIR; T41553; T41553.
 DR HSP; P00572; ITMK.
 DR GeneDB SPombe; SPCC70.07c; -;
 DR InterPro; IPR000062; Thymidylate_kin.
 DR Pfam; PF02223; Thymidylate_kin; 1.
 DR TIGRFAMS; TIGR00041; dTMP_kinase; 1.
 DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
 KW ATP-binding; Kinase; Nucleotide biosynthesis; Transferase.
 FT NP_BIND 14 21 ATP (Potential).
 FT CONFLICT 33 39 SQHEKAE -> LNMKRLK (in Ref. 1).
 FT CONFLICT 59 59 K -> T (in Ref. 1).
 FT CONFLICT 80 93 TIQIVYEQINKGV -> PSIIYRANQQRN (in Ref. 1).
 FT CONFLICT 125 125 P -> T (in Ref. 1).
 FT CONFLICT 164 164 F -> L (in Ref. 1).
 FT CONFLICT 186 186 S -> YA (in Ref. 1).
 FT CONFLICT 191 191 H -> D (in Ref. 1).
 SQ SEQUENCE 210 AA; 24249 MW; 4266144AEDAB68C0 CRC64;
 Query Match 6.1%; Score 144.5; DB 1; Length 210;
 Best Local Similarity 26.1%; Pred. No. 0.014;
 Matches 58; Conservative 39; Mismatches 76; Indels 49; Gaps 11;
 QY 248 QKGRFQVVAIEGLDATGKTTVTSQVADSL-----KAVLLKSPSCIGQWRKIFD----- 296
 DB 3 KQNGRLIVIEGLDRSGKSTCCQLLVKLLSQHEKAELEFPDRTTAIGKIDYLVKESV 62
 QY 297 ----DEPTIIRAFVSLGNYIVASEIAKSAKSPVIVDRVHSTATYATATEVSGGLQHL 352

Db 63 QNDQ---VHLLFSANRWETIQYIEQINKGVTCLDRY----AFSGIAFSAKGLD-- 113
Qy 353 PPAHHEVYQW---PEDLLKPDILILLTVSPEERLQRLQGRGMKTRTEAELEANSVFRQ 408
Db 114 -----WEMCKSPDRGLPRPDLVIFLNVDP--RIAATRGQYGEERYEXIEMQ-----E 158
Qy 409 KVEMSYORME----NPGCHVV--DASPSREKVLQTVLSLIQN 444
Db 159 KVLKNFQRLQKPEFEEGLEFIFITLDASSSLEDVHSQIVDLVSN 200

Search completed: October 22, 2004, 23:28:54
Job time : 163 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2004, 23:32:53 ; Search time 104 Seconds
(without alignments)
1397.766 Million cell updates/sec

Title: US-10-681-223-2

Perfect score: 2362

Sequence: 1 MAFARLLRGPLSGPLGRR.....SRKVLQTVLSLQNSFSEP 449

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1364641 seqs, 323758627 residues

Total number of hits satisfying chosen parameters: 1364641

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US10D_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2362	100.0	449	14	US-10-277-032-2
2	2362	100.0	449	15	US-10-681-223-2
3	795.5	33.7	508	14	US-10-277-032-4
4	795.5	33.7	508	15	US-10-681-223-4
5	155	6.6	193	14	US-10-369-493-911
6	140	5.9	246	15	US-10-282-122A-67367
7	131.5	5.6	205	15	US-10-282-122A-53917
8	128.5	5.4	199	14	US-10-369-493-11256
9	128.5	5.4	253	16	US-10-437-963-196428
10	127.5	5.4	205	14	US-10-369-493-21577
11	127	5.4	259	14	US-10-369-493-22800
12	126.5	5.4	263	15	US-10-425-114-64336
13	125	5.3	210	9	US-09-815-242-11027

11	126	5.3	210	15	US-10-282-122A-58176	Sequence 58176, A
15	125.5	5.3	205	14	US-10-369-493-1320	Sequence 1320, Ap
16	125.5	5.3	205	14	US-10-369-493-20385	Sequence 20385, A
17	125	5.3	212	15	US-10-282-122A-61591	Sequence 61591, A
18	124.5	5.3	212	15	US-10-282-122A-77370	Sequence 77370, A
19	122.5	5.2	1664	16	US-10-437-963-107768	Sequence 107768, A
20	121	5.1	980	14	US-10-156-761-11338	Sequence 11338, A
21	120	5.1	206	15	US-10-282-122A-50167	Sequence 50167, A
22	120	5.1	212	14	US-10-369-493-23045	Sequence 23045, A
23	118.5	5.0	188	14	US-10-369-493-21453	Sequence 21453, A
24	118.5	5.0	216	14	US-10-369-493-22188	Sequence 22188, A
25	118.5	5.0	216	16	US-10-754-929-3	Sequence 3, Appli
26	118.5	5.0	1357	16	US-10-437-963-164009	Sequence 164009, A
27	118	5.0	193	14	US-10-369-493-18097	Sequence 18097, A
28	117.5	5.0	195	14	US-10-369-493-70	Sequence 70, Appl
29	117.5	5.0	213	9	US-09-815-242-10129	Sequence 10129, A
30	117.5	5.0	213	14	US-10-369-493-797	Sequence 797, App
31	117.5	5.0	213	15	US-10-282-122A-56520	Sequence 56520, A
32	117.5	5.0	213	16	US-10-754-929-2	Sequence 2, Appli
33	117.5	5.0	1750	16	US-10-437-963-194022	Sequence 194022, A
34	117	5.0	211	14	US-10-369-493-18331	Sequence 18331, A
35	116.5	4.9	213	15	US-10-282-122A-59458	Sequence 59458, A
36	116.5	4.9	218	14	US-10-369-493-5538	Sequence 5538, Ap
37	116	4.9	196	14	US-10-369-493-8559	Sequence 8559, Ap
38	115	4.9	206	15	US-10-282-122A-47929	Sequence 47929, A
39	114	4.8	1532	16	US-10-437-963-112468	Sequence 112468, A
40	113.5	4.8	210	14	US-10-369-493-286	Sequence 286, App
41	113	4.8	348	14	US-10-369-493-4190	Sequence 4190, Ap
42	112.5	4.8	201	14	US-10-369-493-4374	Sequence 4374, Ap
43	112.5	4.8	206	15	US-10-282-122A-49192	Sequence 49192, A
44	112.5	4.8	213	15	US-10-282-122A-55481	Sequence 55481, A
45	112.5	4.8	1753	16	US-10-437-963-107684	Sequence 107684, A

ALIGNMENTS

RESULT 1

US-10-277-032-2
; Sequence 2, Application US/10277032
; Publication No. US20030087294A1
; GENERAL INFORMATION:
; APPLICANT: Ming-Hui WEI
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001305 DIV
; CURRENT APPLICATION NUMBER: US/10/277,032
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/984,880
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 449
; TYPE: PRT
; ORGANISM: HomoSapien
US-10-277-032-2

Query Match	100.0%	Score	2362	DB	14	Length	449
Best Local Similarity	100.0%	Pred. No.	4.4e+201				
Matches	449	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MAFARLLRGPLSGPLGRRGVGCAGAMAPPCRFVLELPDCTLAHFALGADADADADADP	60				
Db	1	MAFARLLRGPLSGPLGRRGVGCAGAMAPPCRFVLELPDCTLAHFALGADADADADADP	60				
Qy	61	RLAALIGPPERSYSLCPVPTPDAGCGARVRAARLHQLRRGPFQRCOLLRLICYCP	120				
Db	61	RLAALIGPPERSYSLCPVPTPDAGCGARVRAARLHQLRRGPFQRCOLLRLICYCP	120				
Qy	121	GGAGAGAGGFLRRDPLDPTDTEQALLETLLGACQAPRHILGFEADPRGQLWRLEWVQ	180				


```
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CUC001305 DIV-II
; CURRENT APPLICATION NUMBER: US/10/681,223
; CURRENT FILING DATE: 2003-10-09
; PRIOR APPLICATION NUMBER: 10/277,032
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/984,880
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-681-223-4

Query Match 33.7%; Score 795.5; DB 15; Length 508;
Best Local Similarity 48.1%; Pred. No. 9.8e-62;
Matches 177; Conservative 0; Mismatches 4; Indels 187; Gaps 2;

QY 225 ECTSFIPEARAVLDLDVDCPKQKGFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP 284
Db 1 ECTSFIPEARAVLDLDVDCPKQKGFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP 60
QY 285 -----
Db 61 CTSFIPEARAVLDLDVDCPKQKGFQVVAIEGLDATGKTTTQCTSFTEPEARAVLDLDVDC 120
QY 285 -----
Db 121 PKEVQKGFQVIAIEGLDATGKTTLTQHFKSLSLSSYRHPSCIGQWRKIFDDEPTIIR 180
QY 304 RAYFSLGNYIVASEIAKESAKS -----
Db 181 RAYFSLGNYIVASEIAKESAKSPVIVDRYWHSTATYPCIKFNVYVASEIAKESPVIVDRYW 240
QY 326 -----
Db 241 HSTATYPCIKFVEEDLLMNNLSFEFFILWANYLVASEIAKESNTFNPVIVDRYWHSTA 300
QY 338 TYAIAIEVSGGLQHLPPAHPVYQWPEDLKLPDLILLTVPSPERLQGRGMKTRTEE 397
Db 301 TYAIAIEVSGGLQHLPPAHPVYQWPEDLKLPDLILLTVPSPERLQGRGMKTRTEE 360
QY 398 AELEANSV 405
Db 361 AEAIATEV 368

RESULT 5
US-10-369-493-911
; Sequence 911, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 911
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Archaeoglobus fulgidus
US-10-369-493-911

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67367
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-10-282-122A-67367

Query Match 5.9%; Score 140; DB 15; Length 243;
```

```
Query Match 6.6%; Score 155; DB 14; Length 196;
Best Local Similarity 28.6%; Pred. No. 2.2e-05;
Matches 61; Conservative 37; Mismatches 75; Indels 40; Gaps 10;

QY 254 VVAIEGLDATGKTTVTQSVADSL-----KAVLLKSP-PSGIGQWRKIFDDEPTIIRAFY 307
Db 2 LIAVEGIDGAGKTTIAAYIAELLKEGKYKVKVLEKPGDSKFG--KKIKSSEERLSPEEL 59
QY 308 SLGNYIVASEI-AKES-----AKSPVIVDRYWHSTATYATATEVSGGLQHLPPAHPVY 360
Db 60 EL--FLKDREIDARENILPALQSGYAVVMDRYFESNIAVQSGARGIDARL-----IR 108
QY 361 QWPEDLL-KPDLILLTVPSPERLQGRGMKTRTREAELANSVPRQKVENMSYQRMEN 419
Db 109 ENNEKIAPKPDLTILLDVEPEIALEVRVRKRGKLSPEKLD-----YLRVKRCFLENAD 162
QY 420 PGCHVVDASPSRE-----KVLQTVLSLIONS 445
Db 163 ETTVVVDASKPLEBEVKVEVRKVIESFLNLKNS 195
```

RESULT 6

```
US-10-282-122A-67367
; Sequence 67367, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67367
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-10-282-122A-67367
```

Qy	254	VVAIEGLD	TGKTTVT	OSVADSL	KAVLLKSP	SCIGOWR	KIFIDDEP	TIIRRAFYS	----	308
Dd	2	IIAIEGID	GAGKNTL	VSARKFP	ADVI	GFP-----	RYESI	THAKLAQR	ALYSGMDL	54
Qy	309	-----	-----	LGNIV	VASEITAKES	APVIVDR	YHWHSTAT	YAIATE	VS	347
Dd	55	TDSAYAMA	TLPALDRY	DAKAVLG	RYGVGT	SKV-----	VLLDR	VASNAAY	SAAR-----	102
Qy	348	GLQHLP	PAHPHV	PWOPED	-----	LLKPDIL	ILLTLT	VSPEER	QLQ-----	CRGMEX 393
Dd	103	-----	TRNDAMV	QWQVOE	FEELGL	PVPDDI	HLHTS	FELAA	QAQRAREAT	DSKULDR 156
Qy	394	TRBAEAE	LEANSVF	RQKVEMS	QORMEN	PGCHV--	VDASPS-	-REKVLQT	V	438
b	157	VEDFAC	GOERT-FO	AVERSI	AOCKWSP	WLHVHPDES	PDTVOT	RIOTAL	203	

```

RESULT 8
US-10-369-493-11256
; Sequence 11256, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11256
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
; US-10-369-493-11256

```

Query Match	5.4%;	Score 128.5;	DB 14;	Length 199;
Best Local Similarity	26.0%;	Pred.No.0.0052;		
Matches 58;	Conservative 39;	Mismatches 71;	Indels 55;	Gaps 12
Qy	249	KGKFOVAIEGLDATGKTKTIVTSVADSLKAVLLKPPSCIGOWRKIFDDEPT-----I	30101	
Db	1	RGK--LITLEGIDSGSKTVAEK-----LQKNPE--IKAFKPVFTREPTRGTLTGDA	48	
Qy	302	IRRAPYISLGN-----YIVASEIAKESAK-----SPVIVDVRVHSTATYATAEVSG	347	
Db	49	VEKATQSDTDQFAELFLFTADHAELHLIKPALENGKTVISDRYSDSRAYQ-----	101	
Qy	348	GLQHLPPAHHPVYQWPEDI-----LKPDILILLTVSPERLQLOGRGMEKTRREAELEA	402	
Db	102	GITLTRLNPLU--EMWKOLHRSWTVPDITFLDIRPEISIECGRKGQSKEKLE---	157	
Qy	403	NSVFQKQEMSVQRM--ENPFGCHVV--DASPSREKVLQTVLSLI	442	
Db	158	---FTQGVRAFLKTAADDPREFVVIDASRSPEYIEKEVVKI	197	

RESULT 9
US-10-437-963--396428
; Sequence 196428, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihue
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei


```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11027
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-815-242-11027

Query Match          5.3%; Score 126; DB 9; Length 210;
Best Local Similarity 26.1%; Pred. No. 0.0094;
Matches 57; Conservative 32; Mismatches 91; Indels 38; Gaps 10;

QY 249 KGKQVVAIEGLDGTGKTTVTQSVADSL-----KAVLLKSPSC-----IGQWRKIFDD 297
      ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 2 KGKF--IVIEGLEGAGKSAHQSVVRVILHELGIQDVVFTREPGGTFLAEKRLHLIKHETE 59

QY 298 EPTIIRAFYISLGNIVYVASEIAKESAKSP-----VVDYRWHSTATYAIATEVSGGLQH 351
      ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 60 EPVTDKAEILLML--YAARIQLVENVIKPALMQCKVWVGDRHDMSSQAY-----QGGGRQ 111

QY 352 LPPAHPVYQWPEDDL---KPDILLITVSPERLQROGRMEKTRAEAEIANSVFRQ 408
      ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 112 LDP--HFMLTLKXETVLGNFEPDLTIYLDIDPSVGLARARGGELDRIEQMDLD-----FFH 165

QY 409 KVMSYQRM--ENPGCHVVDASPSREKVLQTVLSLION 444
      : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 166 RTRARYLELVKQNPKAVINAEQSIELVQADIESAVKN 203

RESULT 14
US-10-282-122A-58176
; Sequence 58176, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58176

; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 64336
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3632-055-C3_FLI.pep
US-10-425-114-64336

Query Match          5.4%; Score 126.5; DB 15; Length 263;
Best Local Similarity 25.5%; Pred. No. 0.012;
Matches 60; Conservative 34; Mismatches 70; Indels 71; Gaps 13;

QY 250 GKQVVAIEGLDGTGKTTVTQSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFYSL 309
      ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 59 GRGALVVLGELDSGKTSQCARLLSFLKGGYNAE-----GWR--FPDRAT-----SV 104

QY 310 GNVIVASEIAKES-----AKSPVIVDRYWHSTATYAI 342
      ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 105 GQ--MISSYLANDQLDDRTIHLFSANRWEKRALMESKLLSGTTLIVDRYSYS---GVA 159

QY 343 TVESGGIQLPAPHAHPVYQW---PED-LLKPDILLITVSPERLQROGRMEKTRAEAE 398
      ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 160 FSAAKGLD-----IEWCKAPENGLIAPDIVLDVQPEKAAER--GGYGGER--YEKI 208

QY 399 ELEANSVFRQKVMSYQRMENPGCHVVDA-----SPSREKVLQTVLSLIONSFSEP 449
      : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 209 E-----FOKVAEHVHSLRSDTWKVVVDGFLPMETVEEKLRLDLATSCIQQCQNP 257

RESULT 13
US-09-815-242-11027
; Sequence 11027, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```

```
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-10-282-122A-58176

Query Match      5.3%; Score 126; DB 15; Length 210;
Best Local Similarity 26.1%; Pred. No. 0.0094;
Matches 57; Conservative 32; Mismatches 91; Indels 38; Gaps 10;

QY 249 KGKQVVAIEGLDATGKTTVTQSVADSL-----KAVLLKSPSPSC-----IGQWRKIFDD 297
Db 2 KGKF--IVIEGLEGAGKSSAHQSVRVVHELIGIQDVVFTREPGGTPLAEKLRHLIKHETE 59
QY 298 EPTIIRAFYSLGNYIVASEIAKESAKSP-----VIVDRYWHSTATYAIATEVSGGLQH 351
Db 60 EPTVTKRAELML--YAARIQLVENVIKPALMQGKWVVGDRHDMSSQAY-----QGGGRQ 111
QY 352 LPFAHHPVYQWPEDDL---KPDILILLTVSPERLQGRGMEKTRTEAELEANSVFRQ 408
Db 112 LDP--HPMLTKTIVLGNFEFDLTIYLDIDPSVGLARARGELDRIEQMDLD----FFH 165
QY 409 KVMSYQRM--ENPGCHVVDASPSREKVLQTVLSLQN 444
Db 166 RTRARYLELVKDNPKAVVINAQSIQELVQADIESAVKN 203

RESULT 15
US-10-369-493-1320
; Sequence 1320, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkie, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1320
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1320

Query Match      5.3%; Score 125.5; DB 14; Length 205;
Best Local Similarity 23.3%; Pred. No. 0.01;
Matches 53; Conservative 44; Mismatches 71; Indels 59; Gaps 10;

QY 249 KGKQVVAIEGLDATGKTTVTQSVAD-----SLKAVLLKSP-PSCIGQW-RKIFDDEPTI 301
Db 2 RGYF--IVLEGIDGSGKTTQAKLAEWFEDKGYEVLLTKEPTDSELGKLIRRIIEESVI 59
QY 302 I-BRAFYSLGNIVASEIAKESAK-----SPVIVDRYWHSTATYAIATEVSGGLQH 351
Db 60 DGSKISYEAEALLFAADRAEHVKIILPALSEGKVVICDRFYSSL----- 105
QY 352 LPFAHHPVYQWPEDDL-----LKPDLILLTVSPERLQGRGMEKTRTEAE 399
Db 106 -----AYQWARGLDLNLWLIQVNSFAPRPDLAILDLPVKESLRIRKIRG-----TLTE 153
QY 400 LEANSVFRQVMSYQRMEN--PCCHVVDASPSREKVLQTVLSLQN 444
Db 154 FDKIVLQKVRHNYLAEMLFPEMRIVNALSSIEDIHSDIVALVKH 200
```

Search completed: October 22, 2004, 23:40:34
Job time : 107 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2004, 23:05:48 ; Search time 29 Seconds
(without alignments)
1026.786 Million cell updates/sec

Title: US-10-681-223-2

Perfect score: 2362

Sequence: 1 MAFARLLRGPLSGPLGRP.....SREKVLQTVLSLQNSFSEP 449

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2362	100.0	449	4	US-09-984-880-2
2	2362	100.0	449	4	US-10-277-032-2
3	795.5	33.7	508	4	US-09-984-880-4
4	795.5	33.7	508	4	US-10-277-032-4
5	129.5	5.5	936	4	US-09-252-991A-28002
6	118.5	5.0	216	4	US-09-632-553-3
7	118	5.0	691	4	US-09-252-991A-25690
8	117.5	5.0	213	4	US-09-632-553-2
9	117.5	5.0	219	4	US-09-489-039A-7540
10	109	4.6	955	4	US-09-252-991A-24254
11	108.5	4.6	635	4	US-09-252-991A-25642
12	107	4.5	212	3	US-09-259-109-2
13	107	4.5	212	4	US-09-583-110-5141
14	107	4.5	2294	4	US-09-252-991A-17231
15	106	4.5	1427	4	US-09-252-991A-20577
16	105.5	4.5	204	3	US-09-134-001C-3311
17	105.5	4.5	388	4	US-09-252-991A-21572
18	105.5	4.5	564	4	US-09-252-991A-25356
19	105	4.4	444	4	US-09-252-991A-27505
20	104.5	4.4	1665	4	US-09-858-664A-2
21	104.5	4.4	1665	4	US-10-274-978-2
22	104	4.4	494	4	US-09-252-991A-23320
23	104	4.4	1067	4	US-09-252-991A-30526
24	104	4.4	1544	3	US-09-413-814-46
25	103.5	4.4	582	4	US-09-252-991A-25366
26	103.5	4.4	751	4	US-10-020-079-8
27	103.5	4.4	764	4	US-10-020-079-6

```

28 103.5 4.4 864 4 US-10-020-079-4 Sequence 4, Appli
29 103.5 4.4 870 4 US-10-020-079-2 Sequence 2, Appli
30 101.5 4.3 414 4 US-09-252-991A-30034 Sequence 30034, A
31 101.5 4.3 1343 4 US-09-171-991-2 Sequence 2, Appli
32 101 4.3 337 4 US-09-252-991A-31851 Sequence 31851, A
33 100.5 4.3 305 4 US-09-252-991A-20357 Sequence 20357, A
34 100.5 4.3 422 4 US-09-252-991A-30625 Sequence 30625, A
35 100 4.2 280 4 US-09-252-991A-32145 Sequence 32145, A
36 100 4.2 605 4 US-09-252-991A-25512 Sequence 25512, A
37 99.5 4.2 227 4 US-09-543-681A-7769 Sequence 7769, Ap
38 99.5 4.2 435 4 US-09-252-991A-24702 Sequence 24702, A
39 99.5 4.2 507 4 US-09-252-991A-17308 Sequence 17308, A
40 99.5 4.2 657 4 US-09-252-991A-28001 Sequence 28001, A
41 99.5 4.2 776 4 US-10-020-079-24 Sequence 24, Appl
42 99.5 4.2 789 4 US-10-020-079-22 Sequence 22, Appl
43 99.5 4.2 889 4 US-10-020-079-20 Sequence 20, Appl
44 99.5 4.2 895 4 US-10-020-079-18 Sequence 18, Appl
45 99.5 4.2 2285 4 US-09-252-991A-17790 Sequence 17790, A

```

ALIGNMENTS

RESULT 1

US-09-984-880-2

; Sequence 2, Application US/09984880

; Patent No. 6489153

; GENERAL INFORMATION:

; APPLICANT: Ming-Hui WEI

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: C1001305

; CURRENT APPLICATION NUMBER: US/09/984,880

; CURRENT FILING DATE: 2001-10-31

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 449

; TYPE: PRT

; ORGANISM: HomoSapien

; US-09-984-880-2

Query Match 100.0%; Score 2362; DB 4; Length 449;

Best Local Similarity 100.0%; Pred. No. 2.2e-232;

Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 MAFARLLRGPLSGPLGRGVCAGAMAPPCRFVLELPDCTLAHEFALGADAPGDADPDP 60
Db 1 MAFARLLRGPLSGPLGRGVCAGAMAPPCRFVLELPDCTLAHEFALGADAPGDADPDP 60
Qy 61 RLAAALGPERSYSLCVPTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLRLCYCP 120
Db 61 RLAAALGPERSYSLCVPTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLRLCYCP 120
Qy 121 GGAGGAQQGFLRDLDDPDTRQALLELIGACQAPRPHLGFEADPRQQLWQRLWEVQ 180
Db 121 GGAGGAQQGFLRDLDDPDTRQALLELIGACQAPRPHLGFEADPRQQLWQRLWEVQ 180
Qy 181 DGRRLQVGCAQVVPPEPLHPVVPDLPSVVPDPREARAVLEECTSFIPPEARAVLDLV 240
Db 181 DGRRLQVGCAQVVPPEPLHPVVPDLPSVVPDPREARAVLEECTSFIPPEARAVLDLV 240
Qy 241 DQCPKIQKGKFOVAIEGLDATGKTTVTQSVADSLKAVLLKPPSPCIGQWRKIFDDEPT 300
Db 241 DQCPKIQKGKFOVAIEGLDATGKTTVTQSVADSLKAVLLKPPSPCIGQWRKIFDDEPT 300
Qy 301 IIRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAITEVSGGLQHLPPAHHPVY 360
Db 301 IIRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAITEVSGGLQHLPPAHHPVY 360
Qy 361 QWPEDLKPDILILLTVSPEERLORLOGCMKTRFEALEANSVEROKVMSYQRMENP 420

```



```
Db 1 ECTSFIPPEARAVLDVQCPKQIQKGFQVVAIEGLDATGKTTVTSVADSLKAVLLKSP 60
285 -----
Db 61 CTSFIPPEARAVLDVQCPKQKGFQVVAIEGLDATGKTTTQCTSFIPPEARAVLDVQD 120
285 -----
Db 121 PKEVQKGFQVIAIEGLDATGKTTLTQHFKSLRSLSSYRHPSCIGQWRKIFDDEPTIIR 303
285 -----
Db 304 RAYFSLGNYIVASEIAKESAKS 325
285 -----
Db 181 RAYFSLGNYIVASEIAKESAKSPVIVDRYWHSTATPCIKENYVASEIAKESPVIVDRYW 240
326 -----
Db 241 HSTATPPCIKPVEEDLLMNNLSFEPFIFWANYLVASEIAKESNTFPVIVDRYWHSTA 337
338 TYAIAIEVSGLOHLPAPHHVQVQWPEDLAKPDLILLITVSPERLORLOGRGMEKTREE 397
301 TYAIAIEVSGLOHLPAPHHVQVQWPEDLAKPDLILLITVSPERLORLOGRGMEKTREE 360
398 AELEANSV 405
361 AEAIATEV 368

RESULT 5
US-09-252-991A-28002
; Sequence 28002, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28002
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28002

Query Match 5.5%; Score 129.5; DB 4; Length 956;
Best Local Similarity 22.4%; Pred. No. 0.00076;
Matches 124; Conservative 55; Mismatches 182; Indels 193; Gaps 29;

QY 24 AGAMAPPCRF-----VLELPDCTLAHFALGADA-----PGDADAP----- 58
Db 250 AGAVPRLRAADPAALGDLRLQDPDALLHQLAALAPKRPDTCQGPGLDVPRLRHP 309
QY 59 ---DPLA---ALLGPP-----ERSYSLCVPTPDAGCGARVRAARLHORLLHQLRRGP 106
Db 310 RRDAPGLARRPPAQQGPAGGDSRRRRRAVHPAPRPQPPAALAGVRLRRLLHRRRAVD 369
QY 107 FORCOLRLILCYCEGGAG---GAQOGLFLRDLDPDPTQALL-----ELLGACQEA 156
Db 370 PER--LFQQLLPFGFTHRLHAGRGPLRFDP--RPTVQPAAGHAHIRGEHLHACGDS 425
QY 157 PRPHLGEFEADPRGQLWQRLWEVQDGRRLQVGAQVV-----PVPEPPLHPV 203
Db 426 RRARRA---TDPFGLARRFGAA--GKRRAGRGILSLHRLRPLPGYSGGFFRRPQEP 480
QY 204 VPDLPs-----SVVFPDREA-----ARA 221
Db 481 MNDADSLQDYQRVQQAIRSLFEIVGSSSEGTIVVDREARIVWINKRYAQRFGGLADPAQA 540
```

```
QY 222 VLEECTSFIPBEA-----RAVLDVQCPKQIQKGFQVVAIEGLDATGKT--TVT 269
Db 541 IGRPCSEVTPGSLRLREVVGSKPILLDMDSA-----KDFLVVMRLFIHDDAGEVIGAIG 595
QY 270 QSVADSLKAVLLKSP--PSCIGQWRKIFDDEPTI--IRRAFYSYSLGNYIVAS---EIAKES 322
Db 596 FALLDERSI---SPLVTRYLSMRELASTRSOLQARQAKYSFHSFVGTSFASIEVKRRA 652
QY 323 AKSPVIVDRYWHSTAYAIATEVSGGLOHL-----PPAHPVYQ-----WPEDLLKP 369
Db 653 RRGAA-----SEAPVLLLGTTGCKELIAHAIHSASPRAHKAFVSVNAAAIPELSLEA 705
QY 370 DLILLITVSP-----BERLQR-----LQGRGMEKTREAELE 401
Db 706 E---LFGTAPGFTGAERKRPGLQVQAGGTLFLDEIGDMPALQOGK--LLRVLQEKBE 761
QY 402 A---NSVFROKQVEM 412
Db 762 AVGSNEVIRSDIRL 775

RESULT 6
US-09-632-553-3
; Sequence 3, Application US/09632553
; Patent No. 6689595
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn
; TITLE OF INVENTION: Crystallization and Structure Determination of
; TITLE OF INVENTION: Staphylococcus Aureus Thymidylate Kinase
; FILE REFERENCE: 6245.NCP
; CURRENT APPLICATION NUMBER: US/09/632,553
; CURRENT FILING DATE: 2000-08-04
; PRIOR FILING DATE: 60/147,117
; PRIOR APPLICATION NUMBER: 1999-08-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-632-553-3

Query Match 5.0%; Score 118.5; DB 4; Length 216;
Best Local Similarity 25.0%; Pred. No. 0.00097;
Matches 54; Conservative 37; Mismatches 80; Indels 45; Gaps 11;

QY 250 GKFWVAIEGLDATGKTTVTQSVADSLK--AVLLKSP-----SCIGQWRKIFDDEPTI 301
Db 3 GRGKLILIEGLDRTGKTTCQCNILYKLPNCKLLKPERSTRIGGLINEY--LTDDSFQL 60
QY 302 IRAA---FYSYLGNYIVASEIAKESAKSP-VIVDRYWHSTAYAIATEVSGGLOHLPAPHH 357
Db 61 SDQAIHLLFSANRWEIVDKIKKOLLEGKNIVMDRYVYGVAYSAAKGTNG----- 110
QY 358 PVYQW---PE-DLLKXPDILLITVSPERLORLOGRGMEKTREAELEANSVPROKVEMS 413
Db 111 MDLWCLQPDVGLLKPDLTLFLSTQVDNNAEKSGFGDERYE-----TVKFOEVKQT 163
QY 414 YQ-----RMENPGCHVVDASPSREKVLQTVLSLI 442
Db 164 FMKLLDKERKGDSEITIVDVT---NGIQEVEALI 196

RESULT 7
US-09-252-991A-25690
; Sequence 25690, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
```

```
US-09-632-553-2
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25690
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25690

Query Match          5.0%; Score 118; DB 4; Length 681;
Best Local Similarity 23.2%; Pred. No. 0.0067;
Matches 125; Conservative 49; Mismatches 189; Indels 176; Gaps 28;

QY 2 AFARRLLRGLSGPLLRGV-----CAGAMAPPCRFVLEL-----PDCTLAHFALGAD 50
Db 97 AGARRAGPSLGP-AGRRGLRPPPGAGALARPVAVRQVGPARRPPGRLAQ-----RP 151
QY 51 APGDADAPDPLAALLGPPERSYSLCVPTFDAGC--GARVRAARLHORLLHQLRRGPFQ 108
Db 152 QPCTRRRPGDLFPVLRGPKPPRRAGGPDLECRSAENHAALGBERLAASGRPP--- 208
QY 109 RCQILLRLCYCPGGAGGAGGELLRLDPLDDPTQALLLELLGACQAPRPHLGEFADP 168
Db 209 -----APQPGGGRPRG-----RLPLGLLGAVAHADARALLPAAPAE 246
QY 169 R-----GQLWRLWEVQDGRRLQVCAQVVPPEPLHPW--PDLPSVVPFDREAA 220
Db 247 RLHGVSAGPATAGVCAQLQPKR-----GADVPP-----PVDEPGQDPAVHPAAGAR 295
QY 221 AVLEECTSFTEARAVLDLQDCKQIQKGKQFQVAIEGLDAGTKTTTQSVADSLKAVL 280
Db 296 AAI-----AVRDAV--AGHRQHGKPYMT-----LPLSGN-----ALSLEVL 331
QY 281 LKSPSS--C-----IGOWRKIFDDE-PTIIRAFYSLGNIVASEIAKE--SA 323
Db 332 EPIDPGQGFSLFYDPDYDLRLRELREDDSSLETGWQABAKRADWAQVQLASDLQR 391
QY 324 KSPVIVDRYHSTAYATATEVSGLO-----HLPPAHPHVYQ-----W---PE 364
Db 392 RSKDMLAAMUGENAW--LQRGGLGGLQRLALVLAELCERPEEVEHPQADQDSWRVPPI 449
QY 365 D-----LIKPDLLIL-----LTVSPEERLQGRGMKTRBEAELEANSVPRQK 409
Db 450 DWLLRRYAEILLHRLPLMGQGAFAEITLYAWQRLQRCQVASGDSKSAKAALEAAQLOKK 509
QY 410 V-----EMSVORMEN-----PGCHVVDA3PSREKVLQ 436
Db 510 LDEALRAEPLVQWQORKOASLLACQOQLQRLQLEQWCDRCIGELAPSC-----QPLREVIAQ 563

RESULT 8
US-09-632-553-2
; Sequence 2, Application US/09632553
; Patent No. 6689595
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn
; TITLE OF INVENTION: Crystallization and Structure Determination of
; TITLE OF INVENTION: Staphylococcus Aureus Thymidylate Kinase
; FILE REFERENCE: 6245-NCP
; CURRENT APPLICATION NUMBER: US/09/632,553
; CURRENT FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/147,117
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Escherichia coli

US-09-489-039A-7540
; Sequence 7540, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7540
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7540

Query Match          5.0%; Score 117.5; DB 4; Length 219;
Best Local Similarity 22.1%; Pred. No. 0.0013;
Matches 52; Conservative 41; Mismatches 85; Indels 57; Gaps 8;

QY 248 QKGKQFVAIEGLDAGTKTTTQSVADSLKA-----VLLKSPSPSCI----- 288
Db 5 KMRNSNVIVIEGLEGAGKTTARQLVWETLOSAGIHDVFTREPLGLTILAEKLSIVLDIQ 64
QY 289 GOWRKIFDDEPTIIRAFYSLGNIVASEIAKESAKSP-VIVDRYHSTAY----- 339
Db 65 STGDEVINDKAEVL--MFYARVOLVETVVKPALARGOWVIGDRHDJSTQAYQGGRGID 122
QY 340 -----ATATEVSGGLQHLPPAHPVYQWPEDLLKPDILLLTVSPEERLQGRG-MEK 393
Db 123 RTMLATLDAVLGD-----FRPNLTLYLDVTPVGLQARARGELDR 164
QY 394 TREAELEANSVPRQKVMYSORMENPCGHVVDA3PSREKVLQTVLSLIQNSFSE 448
Db 165 IEQESMNFNRTRARYLELA---AADPSIRTVDATQPLDAVARDIRATIAQWMAE 216

RESULT 10
US-09-252-991A-24254
; Sequence 24254, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
```

```
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24254
; LENGTH: 955
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24254

Query Match
Best Local Similarity 28.1%; Pred. No. 0.094;
Matches 74; Conservative 14; Mismatches 83; Indels 92; Gaps 15;

QY 4 ARLLRG-----PLSGPLG-----REGVCA-----GAMAPCRFVLELPDCT 41
DB 706 ARRAVFGAGDPSPVPSRPLLAGPTIDRGLAAGRPAAGDSVQGAAGGTGGGLASDPGAT 765
QY 42 LAHFALGAD---APGDADAPDP-----RLAALLG-----GAPPRSY 73
DB 766 DHGAGLGCGEPGAPGPA-APLPPRQRLMAVGAAGNLGHRGKRTALPPRGTPQPEHRAI 824
QY 74 SLICVPVTPDAGCA-----RVRAARLHQRL---HQLRGPGFORCOLLRLCYCGGQAG 125
DB 825 PRHAP-GPPLGAGAGRRPPRAQPVRLHLRLGGGHCRRQ-----PGQPG 869
QY 126 GAQOQFLLRDPLDDPTQALLEGACQAPRPHLGEFADPRGQLWORLDMEVQDGRLL 185
DB 870 HGPRP-----AQPRPQ-----SPBRPRPRPGEL---PRPTVQVLGQAQRPV 912
QY 186 QVGCAGQWVPVPPPLHPVPPDLP 208
DB 913 RQDRSABLPERQLRP-DGLP 934

RESULT 11
US-09-252-991A-25642
; Sequence 25642, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25642
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25642

Query Match
Best Local Similarity 21.6%; Pred. No. 0.053;
Matches 98; Conservative 49; Mismatches 170; Indels 137; Gaps 20;

QY 6 RLLRGPLSGPLGRRGYCAGAMAPCRFVLELPDCTLAHFALGADAGDADAPDRLAAL 65
DB 280 RLAAGHPGPTADRHMTCKNATLDE---LLEL-----SPRLDADPGVRLALI 325
QY 66 ----LGPERSYSLCVPTPDAGCGARVRAARLHQRLHQLRRGPFQRCOLLRLCYCPG 121
DB 326 ELADLELPEALPLIALRGDPDPGVGEARLLE-----AWEEDAVVDALCAA-- 374
QY 122 GQAGGAQOQFLLRDPLDDPTQALLEGACQE--APR---PHLGEFADPRGQLWRL 176
```

```
DB 375 -----LADPV--PAVADAAQSLGELKEFAAGRLLPLWGHADAFVRASVLRAL 421
QY 177 WEVQDGRRLQVCGAQVVPPEPLHPVDPDPSSVVFFDREARAVLEECTSFIPBARAV 236
DB 422 REL-----RLESAA-----PALAALGDPQAA-----VRREAVAV 451
QY 237 LDLDVQCPKQIQKGFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPSCSI---GOWR 292
DB 452 LGWLRHQPALAEAR---LASADVDFEVRRATGALGSLREATVL--PALCAALADAQWQ 506
QY 293 KIFDDEPTIIRAFYSLGNVIYVASEIAKESASPV---IVDRYWHSTATYAIATEVSGGL 349
DB 507 -----VREEAATLG-----KLGREEAGEPLLKALADDYQVRLR---AARALGRL 549
QY 350 QHLPPAHHPVYQWPEDLKLPDILLITVSPERLORLOGRMEKTRERAELEANSVFRQK 409
DB 550 RHRP-----AREALEALLGHPIGNLRKEAALALGELADPA 584
QY 410 VEMSYQRMENPGCHVVVDASPSREKVLQTVLSLIQ 443
DB 585 SAQALRVABG-----DGDPEVRKAVRIALAQLR 612

RESULT 12
US-09-259-109-2
; Sequence 2, Application US/09259109
; Patent No. 6270762
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K. R.
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Chalker, Allison F.
; APPLICANT: Ingraham, Karen A.
; APPLICANT: Traini, Christopher M.
; APPLICANT: Warren, Patrick V.
; TITLE OF INVENTION: tdk
; FILE REFERENCE: GM10201
; CURRENT APPLICATION NUMBER: US/09/259,109
; CURRENT FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-259-109-2

Query Match
Best Local Similarity 4.5%; Score 107; DB 3; Length 212;
Matches 52; Conservative 34; Mismatches 90; Indels 36; Gaps 8;

QY 254 VVAIEGLDATGKTTVTQSVADSLK---AVLLKSPSCIGQWRKIFD-----DEP 299
DB 6 LVLSLEGAGKTSVLEALLPILKEKGEVLTTRREGVVLGKIREVILDPDSHTQMDAK 65
QY 300 TIIRAFYSLGNVIYVASEIAKESASPVIVDRYWHSTATY-----AIATEVSGGLQHLPP 354
DB 66 TELLYIASRRQHLVKEKVLFALEAGKLVIMDRFIDSVAYQGFGRGLDTEAIDWLN--- 121
QY 355 AHPVVQWPEDLKLPDILLITVSPERLORLOGRMEKTRERAELEANSV-FRQKVEMS 413
DB 122 -----QFATDGLKPLDTLYFDIEVEGLARI---AANSREVNRLDLEGLDHLKKVRQ 172
QY 414 Y-QRMENPGCHVV--DASPSREKVLQTVLSLI 442
DB 173 YLSLLDKENRIVKIDASLPLEQVETTRAVL 204

RESULT 13
US-09-583-110-5141
; Sequence 5141, Application US/09583110
; Patent No. 6699703
```

GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5141
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5141

Query Match 4.5%; Score 107; DB 4; Length 212;
Best Local Similarity 24.3%; Pred. No. 0.014;
Matches 52; Conservative 34; Mismatches 90; Indels 36; Gaps 8;

QY 254 VVAIEGLDATGKTTVTQSVADSLK---AVLLKSPSCIGQWRKIFD-----DEP 299
DB 6 LVSLGEGAGKTSVLEALLPILBEKGVVLTTEPGGVLLGEKIREVILDPHTQNDAX 65
QY 300 TITRAPYISGNVIVASEIAKESAKSPVIVDRVYWHSTATY-----AIATVSGGLQHLPP 354
DB 66 TELLVYASRQHLVEKVLPALEAGKIVIMDRFTDSSVAYQGFGRGIDIEADWLN---- 121
QY 355 AHPVYVQWPEDLKPDLLILLTSPERLQIQGRGNEKTRAEAELEANSV-FRQVEMS 413
DB 122 -----QFATDGKLPDLTLYFDEVEEGLARI---AANSREVNRLDLEGLDLHKVROG 172
QY 414 Y-ORMENPGCHVY--DASPREKVLQTVLSLI 442
DB 173 YLSLKEGNGRIKIDASLEPLEQVVTXAVL 204

RESULT 14
US-09-522-991A-17231
; Sequence 17231, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17231
; LENGTH: 2294
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-522-991A-17231

Query Match 4.5%; Score 107; DB 4; Length 2294;
Best Local Similarity 27.7%; Pred. No. 0.6;
Matches 65; Conservative 16; Mismatches 90; Indels 64; Gaps 11;

QY 5 RLLIRGLSGPLIGRR-----GVCAGAMAPPCRFVLELPDCTTLAHFALGADAPGADA 57
DB 1477 RRLHQGLDQRLGTQPAARPLAGTAGAARARRVAR-----TTA--AVGAPGPGRG 1529
QY 58 PDPRLAALLGPPERSVSLCVPTPDAGC--GARVRAARLHORLLHQLRRGPFORCOLLRL 115

DB 1530 PAPPVAAATAGGA-----AGLRPGQAAEAAGGARLPHQRRREP----- 1557
QY 116 LCYCPGQAGG--AQOGFLLRDLDPDTRQALLLELLGACQEP--RPHLGEFEADPRGQ 171
DB 1568 -----RRAAGVRAQBRQAHRP-----AQRLHLPLRLSRGVPHRQPGQGRHHR--RGT 1613
QY 172 LWORLWEVDGRRLLQVGCQVVPVPEPLHPVVDPLPSSVVPFDRBAARAVLBEC 226
DB 1614 VGARARPERCRRRRQGPRA-----SPVLPGLPAAVGRPARPYRSADHQC 1660

RESULT 15
US-09-252-991A-20577
; Sequence 20577, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20577
; LENGTH: 1427
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20577

Query Match 4.5%; Score 106; DB 4; Length 1427;
Best Local Similarity 22.7%; Pred. No. 0.36;
Matches 113; Conservative 46; Mismatches 189; Indels 150; Gaps 24;

QY 4 ARLLRGLSGPLIGRRGVCAGAMAPPCRFVLELPDCTTLAHFALG-----ADAPGADA 57
DB 773 ARRAREP-----GR-----PADHFDHRAVDLRAAAGAGURDEHLYADRGPHDR 821
QY 58 PD-----PRLAALGP--PERSYSLCVPTPDAGC-----GARVRAARLHORLL 99
DB 822 PDQQAARHPHGRVQRQTATPRAPRPGDPARRADPPAAGADDRGNGLRPGAAALRQR-- 879
QY 100 HQLRRGPFORCOLLRL-----LCYCPGQAGGAGQOGLLRLDPLDD----- 139
DB 880 -RRRQPLRPGRGDRDLRAGRHPLHPVRAHRLYPAGAQRGSRQ-----EPAQPFATGRG 933
QY 140 -----PDTRQALLLEL-----LGACQEP--RP-----HLGEFEADPR 169
DB 934 RSAGEQGMKRSYFNLSRLALALAVGTCLACSVGPDYQRPQSPPPRVASEHLGEFSGERR 993
QY 170 GQLWRLWEVDGRRLLQVGCQVVPVPEPLHPVVDPLPSSVVPFDRBAAR-AVLEBCTS 228
DB 994 EAPW---WSFFDD-----POLVRLVDQALARNHDIRARA 1025
QY 229 FIEBARAVLD--LVDQCPKQIQKGFQVVAIEGIDATGKTTVTSQVADSLKAVLLKSPPS 286
DB 1026 NLESARALFDDRWLDDQLPQVTSQAGYSRSIEQQLDYDGP--RRRLAESYRA-----G 1076
QY 287 CIGWRKIFDEPTIIRRAFYSIGNTIVASEIAKESAKSPVIVDRVYWHSTATVAIATEVS 346
DB 1077 FDAQWE--ID-----LFGRLGRISDAALAEAAADADLRLVLSIAADTARAYFETQ 1126
QY 347 GGLQHLPAPHPHYQWPEDL-LKPDILLITVSPRRRLQRLQGRGMEKTEFEAELE--AN 403
DB 1127 GYQRLDVARAQVRSNRDTLELTRSSLIQSGLPED-----VENAQANLLRSEAAIPPLAT 1182
QY 404 SVFRQKVENSYQRMENPG 421
DB 1183 ALSSARYRLDVLGRGAPG 1200

Search completed: October 22, 2004, 23:33:21
Job time : 32 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2004, 20:24:38 ; Search time 116 Seconds
(without alignments)
1388.530 Million cell updates/sec

Title: US-10-681-223-2

Perfect score: 2362
Sequence: 1 MAFARLLRGLPGILGRR.....SREKVLQTVLSLQNSFSEP 449

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1:	geneseqp1980s:*
2:	geneseqp1990s:*
3:	geneseqp2000s:*
4:	geneseqp2001s:*
5:	geneseqp2002s:*
6:	geneseqp2003as:*
7:	geneseqp2003bs:*
8:	geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2362	100.0	449	6	ABU08678	Novel hum
2	2362	100.0	449	7	ADCE3370	Human nov
3	2362	100.0	449	8	ADN49356	Human kin
4	2350	99.5	449	7	AAE38613	Human thy
5	2350	99.5	449	8	ADI40851	Human kin
6	2350	99.5	449	8	ADP84606	Human bre
7	2177.5	92.2	422	7	ADC39136	Novel hum
8	2177.5	92.2	422	8	ADH42559	Novel hum
9	1552	65.7	395	7	AAE38614	Mouse thy
10	1552	65.7	395	8	ADJ76250	Marker ge
11	1185	50.2	251	7	ADC39138	Novel hum
12	1185	50.2	251	8	ADH42565	Novel hum
13	1137	48.1	290	7	ADF76971	Novel hum
14	982.5	41.6	200	8	ADH42561	Novel hum
15	982.5	41.6	200	8	ADH42563	Novel hum
16	795.5	33.7	508	7	ADCE3372	Mouse thy
17	795.5	33.7	508	8	ADN49358	Mouse thy
18	677.5	28.7	431	5	ABB57297	Mouse isc
19	140	5.9	243	6	ABU39443	Protein e
20	137.5	5.8	283	3	AAG20877	Arabidops
21	132	5.6	208	8	ADN47368	Thermoco
22	131.5	5.6	205	6	ABU25993	Protein e
23	129.5	5.5	956	7	ABO79256	Pseudomon
24	128.5	5.4	224	3	AAG20878	Arabidops
25	128.5	5.4	224	3	AAG50881	Arabidops

26	128.5	5.4	263	3	AAG50880	Arabidops
27	127.5	5.4	205	4	AAB96152	Putative
28	126	5.3	210	2	AAY28787	E.coli th
29	126	5.3	210	4	AAU35434	Haemophil
30	126	5.3	210	6	ABU30252	Protein e
31	126	5.3	210	7	ADJ62328	H.influen
32	125	5.3	212	6	ABU33667	Protein e
33	124.5	5.3	212	6	ABU49446	Protein e
34	124	5.2	227	2	AAR07563	Polypepti
35	123.5	5.2	210	2	AAY28793	Schizosac
36	120	5.1	206	6	ABU22243	Protein e
37	120	5.1	212	2	AAY28785	Thymidyla
38	119.5	5.1	204	2	AAY28794	Thymidyla
39	118.5	5.0	188	2	AAY28789	Thymidyla
40	118.5	5.0	216	2	AAY28796	Teast thy
41	118.5	5.0	216	4	AAE72202	S. cerevi
42	118	5.0	210	7	ADJ62330	H.influen
43	118	5.0	681	7	ABO76944	Pseudomon
44	117.5	5.0	213	2	AAY28786	E.coli th
45	117.5	5.0	213	4	AAU34536	E. coli c

ALIGNMENTS

RESULT 1
ABU08678
ID ABU08678 standard; protein; 449 AA.
XX
AC ABU08678;
XX
DT 10-JUN-2003 (first entry)
XX
DE Novel human thymidylate kinase subfamily kinase.
XX
DE Human; thymidylate kinase subfamily; kinase; therapeutic agent;
KW Human; thymidylate kinase subfamily; kinase; tissue typing;
KW immune response; kinase associated disorder; tissue typing;
KW Pharmacogenomic analysis; enzyme.
XX
OS Homo sapiens.
XX
FH Key
FT Location/Qualifiers
FT Modified-site 17..20
FT /note= "Amidation site"
FT Modified-site 21..26
FT /note= "Myristoylation site"
FT Modified-site 121..126
FT /note= "Myristoylation site"
FT Modified-site 181..184
FT /note= "Amidation site"
FT Modified-site 259..266
FT /note= "ATP/GTP-binding site motif A (P-loop)"
FT Modified-site 259..264
FT /note= "Myristoylation site"
FT Modified-site 263..265
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 271..274
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 275..277
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 322..324
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 378..381
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 390..395
FT /note= "Myristoylation site"
FT Modified-site 394..397
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 407..414
FT /note= "Tyrosine kinase phosphorylation site"
FT Modified-site 445..448
FT /note= "Casein kinase II phosphorylation site"
XX

PN US6489153-B1.
XX 03-DEC-2002.
XX 31-OCT-2001; 2001US-00984880.
XX 31-OCT-2001; 2001US-00984880.
XX (PEKE) PE CORP NY.
XX Wei M, Ketchum KA, Beasley EM, Difrancesco V;
XX WPI; 2003-327246/31.
XX N-PSDB; ABX93989.
XX New human kinase polypeptides and nucleic acid molecules, useful in the
PT development of human therapeutic targets, or for diagnosing or treating a
PT disorder associated with the aberrant expression of the protein or
PT nucleic acid molecule.
XX Claim 1; Fig 2A-C; 49pp; English.
XX The invention describes a new isolated nucleic acid molecule encoding a
CC kinase of the thymidylate kinase subfamily. The human kinase polypeptides
CC and nucleic acid molecules are useful as models for the development of
CC human therapeutic targets, aid in the identification of therapeutic
CC proteins, or serve as targets for the development of human therapeutic
CC agents that modulate kinase activity in cells and tissues that express
CC the kinase. The kinase proteins are also useful in raising antibodies or
CC eliciting another immune response, as reagents in assays designed to
CC quantitatively determine levels of the protein in biological fluids, as
CC markers for tissues in which the corresponding protein is preferentially
CC expressed, or in screening a compound for the ability to stimulate or
CC inhibit interaction between the kinase protein and a molecule that
CC normally interacts with the kinase protein. The kinase proteins and
CC nucleic acid molecules can also be used in providing a target for
CC diagnosing a disease or a predisposition to a disease mediated by the
CC polypeptide or nucleic acid molecule, or for treating a disorder
CC associated with the absence of, inappropriate, or unwanted expression of
CC the protein or nucleic acid molecule. The antibodies are useful for
CC tissue typing, in isolating or detecting kinase proteins, or in
CC pharmacogenomic analysis. This is the amino acid sequence of the novel
CC human thymidylate kinase subfamily kinase
XX Sequence 449 AA;
SQ Query Match 100.0%; Score 2362; DB 6; Length 449;
Best Local Similarity 100.0%; Pred. NO. 1.2e-230;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFARLLRGFLSGPLLRGVCGAGMAPFCRFVLELPDCTLAHFALGADAPGDADPDP 60
DB 1 MAFARLLRGFLSGPLLRGVCGAGMAPFCRFVLELPDCTLAHFALGADAPGDADPDP 60
QY 61 RLAAALGPPRSYSLCVPTVPDAGCGARVRAARLHORLLHQRGPPQRCQLRLLCYCP 120
DB 61 RLAAALGPPRSYSLCVPTVPDAGCGARVRAARLHORLLHQRGPPQRCQLRLLCYCP 120
QY 121 GGQAGGAQQGFLRLDPLDPTQALLELIGACQAPRPHLGEFADPRGOLWORLDV 180
DB 121 GGQAGGAQQGFLRLDPLDPTQALLELIGACQAPRPHLGEFADPRGOLWORLDV 180
QY 181 DGRRLQVCAQVVPPEPLHPVVDLPSSVVFEDREARAVLEECTSFPEARAVLDLV 240
DB 181 DGRRLQVCAQVVPPEPLHPVVDLPSSVVFEDREARAVLEECTSFPEARAVLDLV 240
QY 241 DQCPEKIQKGFQVVALEGLDATGKTTVTSVADSLKXAVLLKSPSCIGQWRKIFDDEPT 300
DB 241 DQCPEKIQKGFQVVALEGLDATGKTTVTSVADSLKXAVLLKSPSCIGQWRKIFDDEPT 300
QY 301 IIRAFYSLGNYIVASIAKESAKSPVIVDRYWEHSTATYATATVSGQLQHLPPAHFPVY 360
DB 301 IIRAFYSLGNYIVASIAKESAKSPVIVDRYWEHSTATYATATVSGQLQHLPPAHFPVY 360

QY 361 QWPEDLLKPDLLILLTVSPERLQRLQGRGMKTRREAELEANSVFRQKVMYSORMENP 420
DB 361 QWPEDLLKPDLLILLTVSPERLQRLQGRGMKTRREAELEANSVFRQKVMYSORMENP 420
QY 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
DB 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
RESULT 2
ADG63370
ID ADG63370 standard; protein; 449 AA.
XX
XX AC ADG63370;
XX 19-DEC-2003 (first entry)
XX Human novel kinase protein.
XX Human; enzyme; kinase; chromosome 2.
XX Homo sapiens.
XX US2003087294-A1.
XX 08-MAY-2003.
XX 22-OCT-2002; 2002US-00277032.
XX 31-OCT-2001; 2001US-00984880.
XX (APPL-) APPLERA CORP.
XX Wei M, Ketchum KA, Beasley EM, Difrancesco V;
XX WPI; 2003-765435/72.
XX N-PSDB; ADG63369, ADG63371.
XX Novel isolated human kinase protein useful for drug screening assays, as
PT a target for diagnosing disease, pharmacogenomic analysis, and for
PT identifying compounds that modulate kinase activity.
XX Claim 1; SEQ ID NO 2; 48pp; English.
XX The invention relates to an isolated peptide consisting of or comprising
CC an amino acid sequence chosen from an amino acid sequence of human kinase
CC protein (ADG63370), an allelic variant of the kinase, orthologue or
CC fragment. Also included are an isolated antibody that selectively binds
CC to the kinase, an isolated nucleic acid molecule encoding the kinase
CC proteins detailed above (appearing as the cDNA, ADG63369, and gene,
CC ADG63371), a gene chip comprising the nucleic acids, a transgenic non-
CC human animal comprising the nucleic acids; a nucleic acid vector
CC comprising the nucleic acids, a host cell containing the vector,
CC detecting the presence of the kinase in a sample, detecting the presence
CC of the nucleic acids in a sample, a pharmaceutical composition comprising
CC a kinase binding agent, treating a disease or condition mediated by a
CC human kinase protein and identifying a modulator of the expression of the
CC kinase involves. The kinase is useful for identifying a modulator, for
CC identifying an agent that binds to the kinase, for biological assays
CC related to kinases, drug screening assays, identifying compounds that
CC modulate kinase activity, to screen a compound for the ability to
CC stimulate or inhibit interaction between the kinase protein and a
CC molecule that normally interacts with kinase protein, for competitive
CC binding assays to discover compounds that interact with kinase,
CC pharmacogenomic analysis and treating a disorder characterised by an
CC absence of inappropriate or unwanted expression of the protein, useful as
CC target spot diagnosing active protein activity, disease or predisposition
CC to disease. The nucleic acid is useful for assessing expression in
CC disease states, diagnostic tools as an immunological marker for aberrant
CC protein for inhibiting protein function, tissue typing. The nucleic acids
CC are useful for facilitating protein trafficking or facilitate
CC manipulation of a protein for assay or production, probes, primers,

CC chemical intermediates, constructing recombinant vectors, expressing
CC antigenic portions of the proteins, for determining chromosomal positions
CC of the nucleic acid molecules, for making vectors containing the gene
CC regulatory regions, as hybridisation probes for determining the presence,
CC level, form and distribution of nucleic acid expression. The gene
CC encoding the kinase is located on human chromosome 2. The present
CC sequence represents the novel kinase.
XX
SQ Sequence 449 AA;
Query Match 100.0%; Score 2362; DB 7; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.2e-230;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFARLLRGPLSGPLGRGVCAGAMAPPCFVLELPDCTLAHFAFGADADGADADP 60
Db 1 MAFARLLRGPLSGPLGRGVCAGAMAPPCFVLELPDCTLAHFAFGADADGADADP 60
QY 61 RLAAALGPPERSYSLCVPTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLRLCYCP 120
Db 61 RLAAALGPPERSYSLCVPTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLRLCYCP 120
QY 121 GGAGGAQGGFLLRDPDPTDTRQALLELLGACQAPRPHLGEFEADPRGQLWQRLWEYQ 180
Db 121 GGAGGAQGGFLLRDPDPTDTRQALLELLGACQAPRPHLGEFEADPRGQLWQRLWEYQ 180
QY 181 DGRRLQVGAQVVPPEPLHPVDPDPSVVPDREAAARVLEECTSFIPPEARAVLDIV 240
Db 181 DGRRLQVGAQVVPPEPLHPVDPDPSVVPDREAAARVLEECTSFIPPEARAVLDIV 240
QY 241 DQCPKQIQKGKQVVAIEGLDGTGKTVTQSVADSLKAVLLKSPSCICQWRKIFDEPT 300
Db 241 DQCPKQIQKGKQVVAIEGLDGTGKTVTQSVADSLKAVLLKSPSCICQWRKIFDEPT 300
QY 301 IIRRAFYSIGNIVASEIAKESAKSPVIVDRYWHSTATVAIATEVSGGLQHLPPAHPYV 360
Db 301 IIRRAFYSIGNIVASEIAKESAKSPVIVDRYWHSTATVAIATEVSGGLQHLPPAHPYV 360
QY 361 QMPEDLLKPDLLILLTVSPERLQIRQGRMGKTRTEAELEANSVFRQKVMYSQRMENP 420
Db 361 QMPEDLLKPDLLILLTVSPERLQIRQGRMGKTRTEAELEANSVFRQKVMYSQRMENP 420
QY 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
Db 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
RESULT 3
ADN49356
ID ADN49356 standard; protein; 449 AA.
XX AC ADN49356;
XX DT 15-JUL-2004 (first entry)
XX DE Human kinase protein.
XX KW Kinase; immune response; protein therapy; human; enzyme.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Modified-site 17..20 /note= "Amidation site"
FT Modified-site 21..26 /note= "N-myristoylation site"
FT Modified-site 121..126 /note= "N-myristoylation site"
FT Modified-site 181..184 /note= "Amidation site"
FT Binding-site 259..266 /note= "ATP/GTP-binding site motif A"
FT Modified-site 259..264

FT Modified-site /note= "N-myristoylation site"
FT 263..285 /note= "protein kinase C phosphorylation site"
FT Modified-site 271..274 /note= "Casein kinase II phosphorylation site"
FT Region 274..294 /note= "Helix 1"
FT Modified-site 275..277 /note= "protein kinase C phosphorylation site"
FT Modified-site 322..324 /note= "protein kinase C phosphorylation site"
FT Modified-site 378..381 /note= "Casein kinase II phosphorylation site"
FT Modified-site 390..395 /note= "N-myristoylation site"
FT Modified-site 394..397 /note= "Casein kinase II phosphorylation site"
FT Modified-site 407..414 /note= "Tyrosine kinase phosphorylation site"
FT Modified-site 445..448 /note= "Casein kinase II phosphorylation site"
XX US2004081999-A1.
XX 29-APR-2004.
XX 09-OCT-2003; 2003US-00681223.
XX 31-OCT-2001; 2001US-00984880.
PR 22-OCT-2002; 2002US-00277032.
XX (APPL-) APPLERA CORP.
XX Wei M, Ketchum KA, Beasley EM, Di Francesco V;
XX WPI: 2004-347669/32.
XX N-PSDB; ADN49355, ADN49357.
XX New isolated human kinase proteins, useful for diagnosing or treating
XX disorders having an absence of, inappropriate, or unwanted expression of
XX the protein.
XX Claim 1; SEQ ID NO 2; 48pp; English.
XX The present invention provides human kinase polypeptides and the encoding
XX polynucleotides. The invention is useful in the development of human
XX therapeutic and diagnostic compositions, to identify therapeutic proteins
XX and serve as targets for the development of human therapeutic agents that
XX modulate kinase activity in cells and tissues, to elicit an immune
XX response and for diagnosing and treating disorders characterised by the
XX absence of inappropriate and unwanted expression of the protein. The
XX polypeptide is also useful in protein therapy. The present sequence is
XX the human kinase protein of the invention.
XX Sequence 449 AA;
Query Match 100.0%; Score 2362; DB 8; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.2e-230;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFARLLRGPLSGPLGRGVCAGAMAPPCFVLELPDCTLAHFAFGADADGADADP 60
Db 1 MAFARLLRGPLSGPLGRGVCAGAMAPPCFVLELPDCTLAHFAFGADADGADADP 60
QY 61 RLAAALGPPERSYSLCVPTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLRLCYCP 120
Db 61 RLAAALGPPERSYSLCVPTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLRLCYCP 120
QY 121 GGAGGAQGGFLLRDPDPTDTRQALLELLGACQAPRPHLGEFEADPRGQLWQRLWEYQ 180
Db 121 GGAGGAQGGFLLRDPDPTDTRQALLELLGACQAPRPHLGEFEADPRGQLWQRLWEYQ 180
QY 181 DGRRLQVGAQVVPPEPLHPVDPDPSVVPDREAAARVLEECTSFIPPEARAVLDIV 240

Db 181 DGRRLQVGCQAVVPEPPHVPVDPDPSVVPDREARAVLEECTSFIPARAVLDLV 240
QY 241 DQCPKQIQKGFQVVAIEGLDGTGTTVTQSVADSLKAVLLKSPSPSCICQWRKIFDEPT 300
Db 241 DQCPKQIQKGFQVVAIEGLDGTGTTVTQSVADSLKAVLLKSPSPSCICQWRKIFDEPT 300
QY 301 IIRAFYSLGNIVASEIAKESAKSPVIVDRYWHSTATVAIATEVSGGLQHLPPAHPVY 360
Db 301 IIRAFYSLGNIVASEIAKESAKSPVIVDRYWHSTATVAIATEVSGGLQHLPPAHPVY 360
QY 361 QWPEDLKPDLLILLTVSPEERLQRLQGRGMKTRREEALEANSVFRQKVMESYQRMENP 420
Db 361 QWPEDLKPDLLILLTVSPEERLQRLQGRGMKTRREEALEANSVFRQKVMESYQRMENP 420
QY 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
Db 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
RESULT 4
AAE38613
ID AAE38613 standard; protein; 449 AA.
AC AAE38613;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human thymidylate kinase protein.
XX
KW Human; thymidylate kinase; EC 2.7.4.9; ATP:dtMP phosphotransferase; TK;
KW enzyme; cancer; gene therapy; chromosome 2p25.2.
XX
OS Homo sapiens.
XX
PN WO2003064642-A1.
XX
PD 07-AUG-2003.
XX
PF 27-JAN-2003; 2003WO-EP000783.
XX
PR 28-JAN-2002; 2002US-0351424P.
PR 25-NOV-2002; 2002US-0428711P.
XX
PA (FARB) BAYER AG.
XX
PI Liou J;
XX
DR WPI; 2003-646151/61.
DR N-PSDB; AAD58542.
XX
PT New polynucleotide and its encoded thymidylate kinase (TK), useful for
PT identifying modulators of TK activity (e.g. agonists or antagonists) that
PT provides therapeutic effects, and in gene therapy for treating cancers.
XX
PS Claim 1; Fig 2; 11opp; English.
XX
XX The invention relates to human thymidylate kinase (TK) polypeptides (EC
CC 2.7.4.9) and polynucleotides. TK is also referred to as ATP:dtMP
CC phosphotransferase. TK sequences are useful for identifying test
CC compounds, that may act as agonists or antagonists at the receptor site
CC and which can be regulated to provide therapeutic effects. Vectors
CC comprising TK polynucleotides are useful in the preparation of a
CC medicament for modulating the activity of TK in a disease, particularly
CC cancer. The invention is also useful in gene therapy. The present
CC sequence is human thymidylate kinase protein. TK gene is located on
CC chromosome 2p25.2
XX
SQ Sequence 449 AA;

Query Match 99.5%; Score 2350; DB 7; Length 449;
Best Local Similarity 99.8%; Pred. No. 2e-229;
Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAFARLLRGPLSGPLIGRRGVCAGAMAPPCRFVLELPDCTTAAHFALGADADADADPDP 60
Db 1 MAFARLLRGPLSGPLIGRRGVCAGAMAPPRFVLELPDCTTAAHFALGADADADADPDP 60
QY 61 RLAAALGPPERSYSLCVPVTPDAGCGARVRAARLHORLLHQLRRGPFQRCQLRLLCYCP 120
Db 61 RLAAALGPPERSYSLCVPVTPDAGCGARVRAARLHORLLHQLRRGPFQRCQLRLLCYCP 120
QY 121 GQAGAGQOQFLLRDPDPTROALLLELGGACQEARPHLGEFEADPRGQMLWRLEWVQ 180
Db 121 GQAGAGQOQFLLRDPDPTROALLLELGGACQEARPHLGEFEADPRGQMLWRLEWVQ 180
QY 181 DGRRLQVGCQAVVPEPPHVPVDPDPSVVPDREARAVLEECTSFIPARAVLDLV 240
Db 181 DGRRLQVGCQAVVPEPPHVPVDPDPSVVPDREARAVLEECTSFIPARAVLDLV 240
QY 241 DQCPKQIQKGFQVVAIEGLDGTGTTVTQSVADSLKAVLLKSPSPSCICQWRKIFDEPT 300
Db 241 DQCPKQIQKGFQVVAIEGLDGTGTTVTQSVADSLKAVLLKSPSPSCICQWRKIFDEPT 300
QY 301 IIRAFYSLGNIVASEIAKESAKSPVIVDRYWHSTATVAIATEVSGGLQHLPPAHPVY 360
Db 301 IIRAFYSLGNIVASEIAKESAKSPVIVDRYWHSTATVAIATEVSGGLQHLPPAHPVY 360
QY 361 QWPEDLKPDLLILLTVSPEERLQRLQGRGMKTRREEALEANSVFRQKVMESYQRMENP 420
Db 361 QWPEDLKPDLLILLTVSPEERLQRLQGRGMKTRREEALEANSVFRQKVMESYQRMENP 420
QY 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
Db 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
RESULT 5
ADI40851
ID ADI40851 standard; protein; 449 AA.
XX
AC ADI40851;
XX
DT 22-APR-2004 (first entry)
XX
DE Human kinase and phosphatase KPP-17 protein SEQ ID NO:17.
XX
KW human; kinase and phosphatase; KPP; enzyme; cardiovascular;
KW antiarteriosclerotic; hypotensive; vasotropic; antiinflammatory;
KW antidiabetic; antiallergic; antidiabetic; nephrotropic; antitumor;
KW gastrointestinal; neuroprotective; osteopathic; antirheumatic; uropathic;
KW ophthalmological; antirheumatic; antiparkinsonian; neurotropic;
KW anticonvulsant; hepatotropic; CNS; antipsoriatic; haemostatic;
KW cytostatic; antilipemic; antiparasitic; antihelminthic; antibacterial;
KW virucide; protozoacide; fungicide; gene therapy; kinase modulator;
KW phosphatase modulator; cardiovascular disease; immune system disorder;
KW neurological disorder; growth and development disorder;
KW cell proliferative disorder; infection.
XX
OS Homo sapiens.
XX
PN WO2004009778-A2.
XX
PD 29-JAN-2004.
XX
PF 18-JUL-2003; 2003WO-US022650.
XX
PR 19-JUL-2002; 2002US-0397354P.
PR 02-AUG-2002; 2002US-0400509P.
PR 02-AUG-2002; 2002US-0400783P.
PR 15-AUG-2002; 2002US-0404027P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Hafalia AJA, Emerling BM, Kable AE, Richardson TW, Becha SD;

PI Baughn MR, Tang YT, Lal PG, Lee SY, Griffin JA, Khare R;
PI Elliott VS, Jin P, Hawkins PK, Swarnakar A, Chawla NK, Tran UK;
PI Gururajan R, Ding L, Marquis JP, Thornton MB, Forsythe JJ, Lee EA;
PI Gietzen KJ, Ramkumar J;
XX
XX
DR WPI; 2004-132950/13.
DR N-PSDB; ADI40304.
XX
XX
PT New human kinases and phosphatases, useful for diagnosing, treating or
PT preventing atherosclerosis, hypertension, AIDS, allergy, multiple
PT sclerosis, osteoarthritis, Alzheimer's disease, Crohn's disease, cancer
PT or hepatitis.
XX
XX
PS Claim 1; SEQ ID NO 17; 330pp; English.
XX
XX
CC The present sequence represents a human kinase and phosphatase (KPP)
CC protein. KPP sequences have cardiovascular, antiarteriosclerotic,
CC hypotensive, vasotropic, antiinflammatory, antiangular, anti-HIV,
CC antiallergic, antiasthmatic, immunosuppressive, antithyroid,
CC dermatological, antidiabetic, nephrotropic, antigout, gastrointestinal,
CC neuroprotective, osteopathic, antiarthritic, uropathic, ophthalmological,
CC antirheumatic, antiparkinsonian, nootropic, anticonvulsant, hepatotropic,
CC CNS, antipsoriatic, haemostatic, cytostatic, antilipaeamic, antiparasitic,
CC antihelminthic, antibacterial, virucide, protozoacide and fungicide
CC activities, and can be used in gene therapy, and as kinase modulators and
CC phosphatase modulators. KPP proteins, polynucleotides, agonists and
CC antagonists can be used for diagnosing, treating or preventing disorders
CC associated with aberrant expression of KPP, such as cardiovascular
CC diseases (e.g. atherosclerosis, hypertension, vasculitis, angina pectoris
CC or congestive heart failure), immune system disorders (e.g. AIDS,
CC allergies, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's
CC disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome,
CC gout, irritable bowel syndrome, multiple sclerosis, osteoarthritis,
CC Sjogren's syndrome or uveitis), neurological disorders (e.g. Alzheimer's
CC disease, Parkinson's disease, Huntington's disease, dementia or
CC epilepsy), disorders affecting growth and development (e.g. cirrhosis,
CC hepatitis, mixed connective tissue disease, psoriasis or primary
CC thrombocytopenia), cell proliferative disorders (e.g.
CC hypercholesterolaemia, hyperlipidaemia or cancer), or viral, bacterial,
CC fungal, parasitic, protozoan or helminthic infections. The KPP and
CC polynucleotides are also useful in assessing the effects of exogenous
CC compounds on the expression of nucleic acids and kinases and
CC phosphatases, or for facilitating the drug discovery process, including
CC determination of efficacy, dosage, toxicity and pharmacology. The
CC polynucleotides encoding KPP are useful for creating transgenic animals
CC to model human disease.
XX
XX
SQ Sequence 449 AA;
Query Match 99.5%; Score 2350; DB 8; Length 449;
Best Local Similarity 99.8%; Pred. No. 2e-229;
Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MAFARLLRGPSGLGRRGVCAGAMAPPCFVLELPDCTLAHFALGADADGADADDP 60
Db 1 MAFARLLRGPSGLGRRGVCAGAMAPPCFVLELPDCTLAHFALGADADGADADDP 60
QY 61 RLAAALGPPERSYSLCVPTPDAGCGARVRAARLHQLRRLHQLRRGPFQRCOLLRLCYCP 120
Db 61 RLAAALGPPERSYSLCVPTPDAGCGARVRAARLHQLRRLHQLRRGPFQRCOLLRLCYCP 120
QY 121 GGAGGAQGGFLRLDPLDDPDTTQALLELLGACQAPRPHLGEFEADPRGQWLWEVQ 180
Db 121 GGAGGAQGGFLRLDPLDDPDTTQALLELLGACQAPRPHLGEFEADPRGQWLWEVQ 180
QY 181 DGRRLVGCAQVVPPEPLHPVVDLPSSVVPDREARAVLEECTSTPEARAVLDIV 240
Db 181 DGRRLVGCAQVVPPEPLHPVVDLPSSVVPDREARAVLEECTSTPEARAVLDIV 240
QY 241 DQCPKIQKGFQVVAIEGLDGTGKTTVTSVADSLKAVLLKSPSCIGOWRKIFDDEPT 300
Db 241 DQCPKIQKGFQVVAIEGLDGTGKTTVTSVADSLKAVLLKSPSCIGOWRKIFDDEPT 300

QY 301 IIRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYATATVSGGLQHLPPAHHPVY 360
Db 301 IIRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYATATVSGGLQHLPPAHHPVY 360
QY 361 QWPEDLLKPDILLILLTVSPERLQRLQGRGMEKTRERAELEANSVFRQKVMYSQRMENP 420
Db 361 QWPEDLLKPDILLILLTVSPERLQRLQGRGMEKTRERAELEANSVFRQKVMYSQRMENP 420
QY 421 GCHVVVDASPSREKVLQTVLSLIQNSFSSEP 449
Db 421 GCHVVVDASPSREKVLQTVLSLIQNSFSSEP 449

RESULT 6

ADP84606
ID ADP84606 standard; protein; 449 AA.

AC ADP84606;
XX
DT 09-SEP-2004 (first entry)
XX
DE Human breast-specific protein #110.
XX
KW human; breast-specific protein; breast cancer.

XX Homo sapiens.

XX OS WO2004053077-A2.

XX PD 24-JUN-2004.

XX PF 05-DEC-2003; 2003WO-US038815.

XX PR 05-DEC-2002; 2002US-0431123P.

XX PA (DIAD-) DIADEXUS INC.

XX PI Macina RA, Turner LR, Sun Y, Chen H, Rodriguez M;

XX DR WPI; 2004-468848/44.

XX DR N-PSDB; ADP84474.

XX
PT New breast specific nucleic acid molecules and polypeptides useful for
PT diagnosing, preventing or treating breast cancer, for producing
PT transgenic animals or cells, or for research purposes.

XX
PS Claim 12; SEQ ID NO 205; 521pp; English.

XX
CC The invention comprises the amino acid and coding sequences of human
CC breast-specific proteins. The DNA and protein sequences of the invention
CC are useful for the diagnosis, treatment and prevention of breast cancer.
CC The present amino acid sequence represents a human breast-specific
CC protein of the invention.

XX
SQ Sequence 449 AA;

Query Match 99.5%; Score 2350; DB 8; Length 449;
Best Local Similarity 99.8%; Pred. No. 2e-229;
Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAFARLLRGPSGLGRRGVCAGAMAPPCFVLELPDCTLAHFALGADADGADADDP 60
Db 1 MAFARLLRGPSGLGRRGVCAGAMAPPCFVLELPDCTLAHFALGADADGADADDP 60
QY 61 RLAAALGPPERSYSLCVPTPDAGCGARVRAARLHQLRRLHQLRRGPFQRCOLLRLCYCP 120
Db 61 RLAAALGPPERSYSLCVPTPDAGCGARVRAARLHQLRRLHQLRRGPFQRCOLLRLCYCP 120
QY 121 GGAGGAQGGFLRLDPLDDPDTTQALLELLGACQAPRPHLGEFEADPRGQWLWEVQ 180
Db 121 GGAGGAQGGFLRLDPLDDPDTTQALLELLGACQAPRPHLGEFEADPRGQWLWEVQ 180

QY 181 DGRRLQVGCAQVVPPEPLHFWPDLPSVYVFPDREARAVLEECTSTFIPARAVLDLV 240
Db 181 DGRRLQVGCAQVVPPEPLHFWPDLPSVYVFPDREARAVLEECTSTFIPARAVLDLV 240
QY 241 DQCPKIQKGKQVVAIEGLDGTGKTTVTQSVADSLKAVLLKSPSPCIGQWRKIFDDEPT 300
Db 241 DQCPKIQKGKQVVAIEGLDGTGKTTVTQSVADSLKAVLLKSPSPCIGQWRKIFDDEPT 300
QY 301 IIRAFYSLGNIVASEIAKESAKSPVIVDRYWHSTATYATIEVSGGLQHLPPAHHPYV 360
Db 301 IIRAFYSLGNIVASEIAKESAKSPVIVDRYWHSTATYATIEVSGGLQHLPPAHHPYV 360
QY 361 QWPEDLKPDLLILLTVSPEERLORLOGRGMKTRAELEANSVFRQVMSYQRMENP 420
Db 361 QWPEDLKPDLLILLTVSPEERLORLOGRGMKTRAELEANSVFRQVMSYQRMENP 420
QY 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
Db 421 GCHVVDASPSREKVLQTVLSLQNSFSEP 449
RESULT 7
ADC39136
ID ADC39136 standard; protein; 422 AA.
XX ADC39136;
XX
XX 18-DRC-2003 (first entry)
DE Novel human NOVX polypeptide SEQ ID NO: 78.
XX
XX antidiabetic; cytostatic; immunomodulator; anorectic; antilipemic;
KW neurotropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
KW antiasthmatic; antiinflammatory; hypotensive; antiarteriosclerotic;
KW hemostatic; osteopathic; gene therapy; NOVX; diabetes; obesity; cancer;
KW lymphoma; uterus cancer; prostate cancer; dyslipidemia; anorexia;
KW wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
KW cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
KW hypertension; atherosclerosis; hemophallia; graft-versus-host disease;
KW Albright hereditary osteodys trophy.
XX
OS Homo sapiens.
XX
XX WO2003010327-A2.
XX
XX 06-FEB-2003.
XX
XX 02-MAY-2002; 2002WO-US014199.
XX
XX 02-MAY-2001; 2001US-0288063P.
XX 03-MAY-2001; 2001US-0288395P.
XX 07-MAY-2001; 2001US-0289087P.
XX 09-MAY-2001; 2001US-0289817P.
XX 09-MAY-2001; 2001US-0289818P.
XX 11-MAY-2001; 2001US-0290194P.
XX 14-MAY-2001; 2001US-0290753P.
XX 15-MAY-2001; 2001US-0291181P.
XX 16-MAY-2001; 2001US-0291243P.
XX 18-MAY-2001; 2001US-0292001P.
XX 21-MAY-2001; 2001US-0292374P.
XX 22-MAY-2001; 2001US-0292587P.
XX 23-MAY-2001; 2001US-0293107P.
XX 25-MAY-2001; 2001US-0293747P.
XX 28-MAY-2001; 2001US-0294109P.
XX 29-MAY-2001; 2001US-0294110P.
XX 30-MAY-2001; 2001US-0294434P.
XX 31-MAY-2001; 2001US-0294827P.
XX 12-JUL-2001; 2001US-0304879P.
XX 31-JUL-2001; 2001US-0308901P.
XX 14-AUG-2001; 2001US-0312270P.
XX 17-AUG-2001; 2001US-0313416P.
XX 10-SEP-2001; 2001US-0318463P.
XX 27-SEP-2001; 2001US-0325683P.
PR

PR 18-OCT-2001; 2001US-0330292P.
PR 28-NOV-2001; 2001US-0338737P.
PR 03-DEC-2001; 2001US-0336909P.
PR 03-DEC-2001; 2001US-0337552P.
PR 21-FEB-2002; 2002US-0359245P.
PR 01-MAY-2002; 2002US-00136826.
XX
PA (CURA-) CURAGEN CORP.
XX Miller CE, Kekuda R, Malyankar UM, Li L, Pena CEA, Spytek KA;
PI Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zerhusen BD;
PI Patturajan M, Anderson DW, Mezes PS, Peyman JA, Macdougall JR;
PI Padigar M, Rastelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;
PI Edinger SR, Ellemann K;
XX
XX WPI; 2003-239445/23.
DR N-PSDB; ADC39135.
XX
XX New NOVX polypeptides and polynucleotides, useful in gene therapy,
PT particularly for treating or preventing a syndrome associated with a
PT human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
PT hypertension or hemophillia.
XX
XX Claim 1; SEQ ID NO 78; 748pp; English.
XX
XX The invention relates to new isolated NOVX polypeptides, the genes
CC encoding them or sequences having at least 95% identity to the amino acid
CC or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,
CC particularly in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, which includes a pathology associated
CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for
CC treating, preventing or alleviating pathology associated with NOVX
CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
CC polypeptide are especially useful for treating or preventing e.g.
CC diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate
CC cancer), dyslipidemia, anorexia, wasting disorders, Alzheimer's disease,
CC Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
CC disease, multiple sclerosis, hypertension, atherosclerosis, hemophillia,
CC graft-versus-host disease or Albright hereditary osteodys trophy. The DNA
CC encoding the protein is useful in gene therapy for treating the above
CC conditions. These are also useful in developing powerful assay system for
CC functional analysis of various human disorders, as well as in diagnostic
CC applications. This sequence represents one of the NOVX proteins of the
CC invention.
XX
XX Sequence 422 AA;
Query Match 92.2%; Score 2177.5; DB 7; Length 422;
Best Local Similarity 99.1%; Pred. No. 6.1e-212;
Matches 420; Conservative 0; Mismatches 1; Indels 3; Gaps 2;
QY 27 MAPPCRFVLELPDCTLAHFALGADADADPPRLAALLGPPERSYSLCVPTPDAGCG 86
Db 1 MAPPCRFVLELPDCTLAHFALGADADADPPRLAALLGPPERSYSLCVPTPDAGCG 60
QY 87 ARVRAARLHORLHQRGPFQFCQLRLCYCPGGAGGAGCGFLRDLDDPDTFOAL 146
Db 61 ARVRAARLHORLHQRGPFQFCQLRLCYCPGGAGGAGCGFLRDLDDPDTFOAL 120
QY 147 LELLGACQAPRPHLGEFEADPRGQLWQRLEWQDGRRLQVGCQVVPPEPLHFWPVD 206
Db 121 LELLGACQAPRPHLGEFEADPRGQLWQRLEWQDGRRLQVGCQVVPPEPLHFWPVD 180
QY 207 LPSSVVPDREARAVLEECTSTFIPARAVLDLVQCPKIQKGKQVVAIEGLDGTGK 265
Db 181 LPSSVVPDREARAVLEECTSTFIPARAVLDLVQCPKIQKGKQVVAIEGLDGTGK 240
QY 266 TTVTQSVADSLKAVLLKSPSPCIGQWRKIFDDEPTIIRAFYSLGNIVASEIAKESAKS 325
Db 241 TTVTQSVADSLKAVLLKSPSPCIGQWRKIFDDEPTIIRAFYSLGNIVASEIAKESAKS 300
QY 326 FVIVDRYWHSTATYATIEVSGGLQHLPPAHHPYQWPEDLKPDLLILLTVSPEERLOR 385

Db 301 PVIDR-HSTATYATATEVSGGLQHLPAHPHYQWPDLLKPDILLITVSPERLQ 358
QY 386 LQGRGMEKTRAEAELEANSVFRQKVMYSQRMENPCGHVVDAAPSREKVLQTVLSIQNS 445
Db 359 LQGRGMEKTRAEAELEANSVFRQKVMYSQRMENPCGHVVDAAPSREKVLQTVLSIQNS 418
QY 446 FSEP 449
Db 419 FSEP 422
RESULT 8
ADH42559
ID ADH42559 standard; protein; 422 AA.
XX AC ADH42559;
XX DT 25-MAR-2004 (first entry)
XX DE Novel human protein NOV78a.
XX KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;
KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;
KW antiparkinsonian; antiaschmatic; antiinfertility; cardiomyopathy;
KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;
KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.
XX OS Homo sapiens.
XX PN WO2003102159-A2.
XX PD 11-DEC-2003.
XX PF 04-JUN-2003; 2003WO-US017573.
PF 04-JUN-2002; 2002US-0385490P.
PR 04-JUN-2002; 2002US-0385615P.
PR 04-JUN-2002; 2002US-0385755P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386355P.
PR 06-JUN-2002; 2002US-0386357P.
PR 06-JUN-2002; 2002US-0386447P.
PR 06-JUN-2002; 2002US-0386459P.
PR 06-JUN-2002; 2002US-0386465P.
PR 07-JUN-2002; 2002US-0386701P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0387078P.
PR 07-JUN-2002; 2002US-0387081P.
PR 10-JUN-2002; 2002US-0387429P.
PR 10-JUN-2002; 2002US-0387429P.
PR 10-JUN-2002; 2002US-0387540P.
PR 10-JUN-2002; 2002US-0387866P.
PR 11-JUN-2002; 2002US-0387606P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387659P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387696P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 12-JUN-2002; 2002US-0388432P.
PR 12-JUN-2002; 2002US-0388479P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389604P.

PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 25-JUN-2002; 2002US-0391726P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403458P.
PR 15-AUG-2002; 2002US-0403617P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 23-SEP-2002; 2002US-0412955P.
PR 30-SEP-2002; 2002US-0415195P.
PR 23-OCT-2002; 2002US-0420627P.
PR 23-OCT-2002; 2002US-0420718P.
PR 24-OCT-2002; 2002US-0420852P.
PR 31-OCT-2002; 2002US-0422750P.
PR 01-NOV-2002; 2002US-0423095P.
PR 05-NOV-2002; 2002US-0423748P.
XX (CURA-) CURAGEN CORP.
XX PA Alsbrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
XX PI Burgess CE, Casman SJ, Catterton E, Dhanabal M, Edinger SR;
PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI Grose NW, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;
PI Khramtsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;
PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;
PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
PI Wolenc AR, Zhong M, Zhong H;
XX WPI: 2004-053467/05.
DR N-PSDB; ADH42558.
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or
XX treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
PT pharmacogenomics.
XX Claim 2; SEQ ID NO 1112; 1503pp; English.
XX The invention relates to 566 new isolated human polypeptides and their
CC encoding genes, sequences that are at least 95% identical to these or
CC sequences comprising one or more conservative substitutions in these. The
CC polypeptide, polynucleotide and antibodies against the polypeptides are
CC useful in diagnosing, treating or preventing NOVX-associated disorders,
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
CC The nucleic acids are further used as hybridization probes, in chromosome
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
CC polypeptides are also useful as vaccines. This sequence represents an
CC example of the polypeptide of the invention.
XX Sequence 422 AA;

Query Match 92.2%; Score 2177.5; DB 8; Length 422;
Best Local Similarity 99.1%; Pred. No. 6.1e-212;
Matches 420; Conservative 0; Mismatches 1; Indels 3; Gaps 2;
QY 27 MAPPCRFVLELPDCTLAHFALGADAPGDADAPDLAALGPPERSYSLCVPVTPDAGCG 86
Db 1 MAPPCRFVLELPDCTLAHFALGADAPGDADAPDLAALGPPERSYSLCVPVTPDAGCG 60
QY 87 ARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCPGGAGGAGGFLRRDLPDPTQAL 146
Db 61 ARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCPGGAGGAGGFLRRDLPDPTQAL 120

QY 147 LELLGACQAPRPHLGEFEADPRGOLWQRLWEVDGRRLOVGCQAVVPVPEPLHPVVD 206
DB 121 LELLGACQAPRPHLGEFEADPRGOLWQRLWEVDGRRLOVGCQAVVPVPEPLHPVVD 180
QY 207 LPSSVVPDREARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKQVVAIEGLDAT-GK 265
DB 181 LPSSVVPDREARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKQVVAIEGLDATGK 240
QY 266 TTVTQSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFYSLGNVIYVASEIAKESAKS 325
DB 241 TTVTQSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFYSLGNVIYVASEIAKESAKS 300
QY 326 PVIIVRYWSTATYATATEVSGLOHLPAAHHPVYQWPEDLKPDILILLTVSPEERLOR 385
DB 301 PVIIVDR--HSTATYATATEVSGLOHLPAAHHPVYQWPEDLKPDILILLTVSPEERLOR 358
QY 386 LQGRGMEKTRAEAELEANSVFRQKVEMSYORMENPGCHVVVDASPSREKVLQTVLSLIONS 445
DB 359 LQGRGMEKTRAEAELEANSVFRQKVEMSYORMENPGCHVVVDASPSREKVLQTVLSLIONS 418
QY 446 FSEP 449
DB 419 FSEP 422

RESULT 9

AAE38614
ID AAE38614 standard; protein; 395 AA.

AC AAE38614;

DT 04-DEC-2003 (first entry)

XX Mouse thymidylate kinase protein.

XX Mouse; thymidylate kinase; EC 2.7.4.9; ATP:dTMP phosphotransferase; TK;
KW enzyme; cancer; gene therapy.

XX Mus musculus.

XX WO2003064642-A1.

XX 07-AUG-2003.

XX 27-JAN-2003; 2003WO-EP000783.

XX 28-JAN-2002; 2002US-0351424P.

XX 25-NOV-2002; 2002US-0428711P.

XX (FARB) BAYER AG.

XX Liou J;

XX WPI; 2003-646151/61.

XX New polynucleotide and its encoded thymidylate kinase (TK), useful for
PT identifying modulators of TK activity (e.g. agonists or antagonists) that
PT provides therapeutic effects, and in gene therapy for treating cancers.

XX Disclosure; Page 88-89; 110pp; English.

XX The invention relates to human thymidylate kinase (TK) polypeptides (EC
CC 2.7.4.9) and polynucleotides. TK is also referred to as ATP:dTMP
CC phosphotransferase. TK sequences are useful for identifying test
CC compounds, that may act as agonists or antagonists at the receptor site
CC and which can be regulated to provide therapeutic effects. Vectors
CC comprising TK polynucleotides are useful in the preparation of a
CC medicament for modulating the activity of TK in a disease, particularly
CC cancer. The invention is also useful in gene therapy. The present
CC sequence is mouse thymidylate kinase protein. This sequence is used in
CC the invention

XX Sequence 395 AA;

Query Match 65.7%; Score 1552; DB 7; Length 395;
Best Local Similarity 77.8%; Pred. No. 2.2e-148;
Matches 294; Conservative 36; Mismatches 48; Indels 0; Gaps 0;

QY 68 PERSYSVLCVPTPDAGCGARVRAARLHORLHQLRRGPFQRCQLLELLCYCPGGQAGGA 127
DB 16 PQGAATRLCVPLAPGEGCGGPRVQAAHVHRLQLQRRGFLQRCQLSKLLGYGPGDQAGEA 75
QY 128 QQGFLLRDLDDPDTRQALLELLGACQAPRPHLGEFEADPRGOLWQRLWEVDGRRLOV 187
DB 76 QHGFLLRDLDDPDTRDILLQLGSCQEAARFQAEFFQADSQGLLWQRLWEVDGRRLOV 135
QY 188 GCAQVVPVPEPLHPVVPDLPSSVVPDREARAVLEECTSFIPEARAVLDLVDQCPKQI 247
DB 136 DCACVLPAPGPHLHLLPDLNLNSAVFQDRDARAVLEECTSFIPEARAVLDLVDQCPKEV 195
QY 248 QKGKFOVVAIEGLDATGKTTVTQSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFY 307
DB 196 QKGKFOVVAIEGLDATGKTTLTQSVSESLKAVLLQSPFPCISQWRKIFDDEPTIIRAFY 255
QY 308 SLGNYIVASEIAKESAKSPVIIVDRYWHSTATYATATEVSGLOHLPAAHHPVYQWPEDL 367
DB 256 SLGNYIVASEIAKESAKSTNFPVIVDRYWHSTATYATATEVSGLOHLPAAHHPVYQWPEDL 315
QY 368 KPDILILLTVSPEERLORLQGRGMEKTRAEAELEANSVFRQKVEMSYORMENPGCHVVDA 427
DB 316 KPDILVLLLTWNSERVRRLQGRGQEKTKBEAELEANNVFRQKVEMTYORMENPCHLVDA 375
QY 428 SPSREKVLQTVLSLIONS 445
DB 376 SPSRETVLQKVLLELIQSS 393

RESULT 10

ADJ76250

ID ADJ76250 standard; protein; 395 AA.

XX AC ADJ76250;

XX 20-MAY-2004 (first entry)

XX Marker gene related amino acid sequence SEQ ID NO:1502.

XX bronchial asthma; chronic obstructive pulmonary disease;
KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;
KW gene therapy; marker.

XX Mus musculus.

XX EP1394274-A2.

XX 03-MAR-2004.

XX 04-AUG-2003; 2003EP-00254857.

XX 06-AUG-2002; 2002JP-00229312.

XX 20-MAR-2003; 2003JP-00077212.

XX (GENO-) GENOX RES INC.

XX Ontani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;

XX WPI; 2004-193155/19.

XX Testing for bronchial asthma or chronic obstructive pulmonary disease by
PT comparing the expression level of a marker gene in a biological sample
PT from a subject with the expression level of the gene in a sample from a
PT healthy subject.

XX Claim 16; SEQ ID NO 1502; 241pp; English.

XX The present invention describes a method of testing for bronchial asthma

or chronic obstructive pulmonary disease. The method comprises determining the expression level of a marker gene in a biological sample from a subject, comparing the expression level determined with the expression level of the marker gene in a biological sample from a healthy subject, and judging whether the subject has bronchial asthma or chronic obstructive pulmonary disease. The marker gene comprises: (a) a group of genes (S1) whose expression levels increase when respiratory epithelial cells are stimulated with interleukin-13; or (b) a group of genes (S2) whose expression levels decrease when respiratory epithelial cells are stimulated with interleukin-13. Also described: (1) a reagent (I) for testing for bronchial asthma or chronic obstructive pulmonary disease; (2) a kit for screening for a candidate compound for a therapeutic agent to treat bronchial asthma or chronic obstructive pulmonary disease; (3) an animal model for bronchial asthma or chronic obstructive pulmonary disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a method for producing an animal model for bronchial asthma or chronic obstructive pulmonary disease; (6) a therapeutic agent for bronchial asthma or chronic obstructive pulmonary disease, comprising the compound, a marker gene or an antisense nucleic acid corresponding to a portion of the marker gene, a ribozyme, a polynucleotide that suppresses the expression of the gene through an RNAi effect or an antibody recognising a protein encoded by a marker gene; and (7) a DNA chip for testing for bronchial asthma or a chronic obstructive pulmonary disease, on which a probe has been immobilised to assay a marker gene. (I) has respiratory and antiasthmatic activities, and can be used in gene therapy. The method is useful for testing for or screening for a therapeutic agent for bronchial asthma or chronic obstructive pulmonary disease. The present sequence is used in the exemplification of the present invention.

XX
SQ Sequence 395 AA;

Query Match 65.7%; Score 1552; DB 8; Length 395;
Best Local Similarity 77.8%; Pred. No. 2.2e-148;
Matches 294; Conservative 36; Mismatches 48; Indels 0; Gaps 0;

QY 68 PPERSYSLCVPTDAGCGARVRAARLHQLRGGPFCORCOLRLLCYCPGGAGGA 127
DB 16 PQGATRLCVLAPGEGCGPRVQAAVHRLQLQRRGLRQLSKLLGYGEGDQAGEA 75

QY 128 QGFLRLDPLDPTDTRQALLGLGACQAPRHLGFEADPRGQLWQRLWEVQDGRRLQV 187
DB 76 QHGFLLRDPCHDPTDRDLQLLGLSCQEAARPOLAEFQADSQGLLWQRLWEVQDGRVQV 135

QY 188 GCAQVVPVPEPLHPVVPDLPSVVPDREARAVLEECTSFIPPEARVLDLVDQCPKQI 247
DB 136 DCACVLPAGEPHLHPLDPLLSAVFQDRDAARAVLEECTSFIPPEARVLDLVDQCPKEV 195

QY 248 QKGKFOVAIEGLDAGTKTQVTSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFY 307
DB 196 QKGKFOVAIEGLDAGTKTQVTSVSESLKAVLLQSPFPCISQWRKIFDDEPTIIRAFY 255

QY 308 SLGNIVASEIAKESAKSPVIVDRYVHSTATYAITEVSGGLQHLPPAHPHYVQWPEDL 367
DB 256 SLGNIVASEIAKESSTNFPVIVDRYVHSTATYAITEVSGGLQYLPAPHPHYVQWPGDLL 315

QY 368 KPDILLITVSPERLCRQGRMEKTRAEAELEANSVFRQKVMYSQRMENPCGHVDA 427
DB 316 KPDVLLITVNSERVRLQGRQGEKTKBEAELEANNVFRQKVMYSQRMENPNSCHLVDA 375

QY 428 SPSREKVLQTVLSLIONS 445
DB 376 SPSRETVLQVLELIQSS 393

RESULT 11
ADC39138
ID ADC39138 standard; protein; 251 AA.
XX ADC39138;
XX
XX 18-DEC-2003 (first entry)
XX Novel human NOVX polypeptide SEQ ID NO: 80.
DE

XX
KW antidiabetic; cytostatic; immunomodulator; anorectic; antilipemic;
KW neurotropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
KW antiasthmatic; antiinflammatory; hypotensive; antiarteriosclerotic;
KW hemostatic; osteopathic; gene therapy; NOVX; diabetes; obesity; cancer;
KW lymphoma; uterine cancer; prostate cancer; dyslipidemia; anorexia;
KW wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
KW cardiomyopathy; AIDS; asthma; Cronh's disease; multiple sclerosis;
KW hypertension; atherosclerosis; hemophilia; graft-versus-host disease;
KW Albright hereditary osteodystrophy.
XX
OS Homo sapiens.
XX
PN WO2003010327-A2.
XX
PD 06-FEB-2003.
XX
PF 02-MAY-2002; 2002WO-US014199.
XX
PR 02-MAY-2001; 2001US-0288063P.
PR 03-MAY-2001; 2001US-0288395P.
PR 07-MAY-2001; 2001US-0289087P.
PR 09-MAY-2001; 2001US-0289817P.
PR 11-MAY-2001; 2001US-0289818P.
PR 14-MAY-2001; 2001US-0290194P.
PR 15-MAY-2001; 2001US-0290753P.
PR 16-MAY-2001; 2001US-0291181P.
PR 18-MAY-2001; 2001US-0291243P.
PR 21-MAY-2001; 2001US-0292001P.
PR 22-MAY-2001; 2001US-0292374P.
PR 23-MAY-2001; 2001US-0292587P.
PR 25-MAY-2001; 2001US-0293107P.
PR 29-MAY-2001; 2001US-0293747P.
PR 29-MAY-2001; 2001US-0294109P.
PR 30-MAY-2001; 2001US-0294110P.
PR 31-MAY-2001; 2001US-0294434P.
PR 12-JUL-2001; 2001US-0294827P.
PR 31-JUL-2001; 2001US-0304879P.
PR 14-AUG-2001; 2001US-0308901P.
PR 17-AUG-2001; 2001US-0312270P.
PR 10-SEP-2001; 2001US-0313416P.
PR 27-SEP-2001; 2001US-0318463P.
PR 18-OCT-2001; 2001US-0325683P.
PR 28-NOV-2001; 2001US-0330292P.
PR 03-DEC-2001; 2001US-0333873P.
PR 03-DEC-2001; 2001US-0336909P.
PR 21-FEB-2002; 2001US-0337552P.
PR 01-MAY-2002; 2002US-0359245P.
XX
XX (CURA-) CURAGEN CORP.
XX
PI Miller CE, Kekuda R, Malyankar UM, Li L, Pena CEA, Spytek KA;
PI Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zerhusen BD;
PI Patturajan M, Anderson DW, Mezes PS, Peyman JA, Macdougall JR;
PI Padgugaru M, Rastelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;
PI Edinger SR, Eilerman K;
XX
DR WPI; 2003-239445/23.
DR N-PSDB; ADC39137.
XX
XX
PT New NOVX polypeptides and polynucleotides, useful in gene therapy, a particularly for treating or preventing a syndrome associated with a human disease e.g. diabetes, obesity, cancer, Alzheimer's disease, hypertension or hemophilia.
XX
PS Claim 1; SEQ ID NO 80; 748pp; English.
XX
CC The invention relates to new isolated NOVX polypeptides, the genes encoding them or sequences having at least 95% identity to the amino acid or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic, particularly in the manufacture of a medicament for treating a syndrome associated with a human disease, which includes a pathology associated

CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for
CC treating, preventing or alleviating pathology associated with NOVX
CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
CC polypeptide are especially useful for treating or preventing e.g.
CC diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate
CC cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease,
CC Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
CC disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,
CC graft-versus-host disease or Albritght hereditary osteodystrophy. The DNA
CC encoding the protein is useful in gene therapy for treating the above
CC conditions. These are also useful in developing powerful assay system for
CC functional analysis of various human disorders, as well as in diagnostic
CC applications. This sequence represents one of the NOVX proteins of the
CC invention.

XX SQ Sequence 251 AA;
Query Match 50.2%; Score 1185; DB 7; Length 251;
Best Local Similarity 58.9%; Pred. No. 2.2e-111;
Matches 249; Conservative 0; Mismatches 2; Indels 172; Gaps 1;
QY 27 MAPPCRFVLELPDCTLAHFAALGADADPDRLAALLGPPERSYSLCVPTPDAGCG 86
Db 1 MAPPCRFVLELPDCTLAHFAALG----- 22
QY 87 ARVRAARLHORLLHQLRRGPFQRCQLLLRLCYCGGAGGAQGFLLRDLDDPTFQAL 146
Db 23 ----- 22
QY 147 LELLGACQAPRPHLGEFEADPRQLWRLWEVDGRRLLQVCAQVVPVPEPLHPVVPD 206
Db 23 ----- 22
QY 207 LPSSVVPDREARAVLEBCTSFIPERAVLDLDVQCPKIQKGFQVVAIEGLDGTGT 266
Db 23 -----AVLEBCTSFIPERAVLDLDVQCPKIQKGFQVVAIEGLDGTGT 68
QY 267 TWTQSVADSLKAVLLKSPSCIGQWRKIFDEPTITPRAFVSLGNIVASEIAKESAKSP 326
Db 69 TWTQSAADSLKAVLLKSPSCIGQWRKIFDEPTITPRAFVSLGNIVASEIAKESAKSP 128
QY 327 VIIVDRVHSTATYATATEVSGGLQHLPPAHPVYQWPEDLKPDLLILLTVSPERLQRL 386
Db 129 VIIVDRVHSTATYATATEVSGGLQHLPPAHPVYQWPEDLKPDLLILLTVSPERLQRL 188
QY 387 QGRGMKTRBEAELEANSVPQKVENSYQRMENPGCHVVDASPSREKVLQTVLSLQNSF 446
Db 189 QGRGMKTRBEAELEANSVPQKVENSYQRMENPGCHVVDASPSREKVLQTVLSLQNSF 248
QY 447 SEP 449
Db 249 SEP 251

RESULT 12
ADH42565
ID ADH42565 standard; protein; 251 AA.
AC ADH42565;
XX
XX
DT 25-MAR-2004 (first entry)
XX
DE Novel human protein NOV78d.
XX
KW cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;
KW antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;
KW antiparkinsonian; antiasthmatic; antiinfertility; cardiomyopathy;
KW atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;
KW multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
KW tissue typing; preventive medicine; pharmacogenomic; vaccine.
XX Homo sapiens.
OS

XX WO2003102159-A2.
PN
XX
XX 11-DEC-2003.
XX
PF 04-JUN-2003; 2003WO-US017573.
XX
XX 04-JUN-2002; 2002US-0385490P.
PR 04-JUN-2002; 2002US-0385615P.
PR 04-JUN-2002; 2002US-0385755P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386355P.
PR 06-JUN-2002; 2002US-0386357P.
PR 06-JUN-2002; 2002US-0386447P.
PR 06-JUN-2002; 2002US-0386459P.
PR 06-JUN-2002; 2002US-0386465P.
PR 06-JUN-2002; 2002US-0386864P.
PR 07-JUN-2002; 2002US-0386701P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0387078P.
PR 07-JUN-2002; 2002US-0387081P.
PR 07-JUN-2002; 2002US-0387083P.
PR 10-JUN-2002; 2002US-0387429P.
PR 10-JUN-2002; 2002US-0387540P.
PR 10-JUN-2002; 2002US-0387866P.
PR 11-JUN-2002; 2002US-0387606P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387659P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387696P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 12-JUN-2002; 2002US-0388432P.
PR 12-JUN-2002; 2002US-0388479P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389604P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 25-JUN-2002; 2002US-0391726P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403458P.
PR 15-AUG-2002; 2002US-0403177P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 30-SEP-2002; 2002US-0412955P.
PR 23-OCT-2002; 2002US-0420627P.
PR 23-OCT-2002; 2002US-0420718P.
PR 24-OCT-2002; 2002US-0420852P.
PR 31-OCT-2002; 2002US-0422750P.
PR 01-NOV-2002; 2002US-0423095P.
PR 03-NOV-2002; 2002US-0423748P.

(CURA-) CURAGEN CORP.
XX
XX Alsbrook CE, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
PI Burgess CP, Casman SJ, Catterton B, Dhanabal M, Edinger SR;
PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI Grosse WW, Gunther E, Guo X, Gusev YI, Herrmann JL, Ji W, Kekuda R;
PI Khramtsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;

PI Maclachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;
 PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;
 PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;
 PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
 PI Wolenc AR, Zhong M, Zhong H;
 XX WPI: 2004-053467/05.
 DR N-PSDB; ADH42564.
 XX New NOVX polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
 PT pharmacogenomics.
 XX
 PS Claim 2; SEQ ID NO 118; 1503pp; English.
 XX
 CC The invention relates to 566 new isolated human polypeptides and their
 CC encoding genes, sequences that are at least 95% identical to these or
 CC sequences comprising one or more conservative substitutions in these. The
 CC polypeptide, polynucleotide and antibodies against the polypeptides are
 CC useful in diagnosing, treating or preventing NOVX-associated disorders,
 CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
 CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
 CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
 CC The nucleic acids are further used as hybridization probes, in chromosome
 CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
 CC polypeptides are also useful as vaccines. This sequence represents an
 CC example of the polypeptide of the invention.
 XX
 SQ Sequence 251 AA;

Query Match 50.2%; Score 1185; DB 8; Length 251;
 Best Local Similarity 58.9%; Pred. No. 2.2e-111;
 Matches 249; Conservative 0; Mismatches 2; Indels 172; Gaps 1;
 QY 27 MAPPCRFVLELPDCTLAHFAFGADPGDADDPRLAALLGPPERSYSLCVPTPDAGCG 86
 DB 1 MAPPPRFVLELPDCTLAHFAFG----- 22
 QY 87 ARVRAARLHQRLHQLRRGFFORCQLRLCYCGGAGGAGQGLRLDPLDPTDTRQAL 146
 DB 23 ----- 22
 QY 147 LELLGACQAPRHLGEFEADPRGQLWQRLWEVDGRRLQVCAQVVPVPEPLHPVVD 206
 DB 23 ----- 22
 QY 207 LPSSVVPDREARAVLECTSPIPEARAVLDLDVQCCKQIOGKGFQWVAIEGLDATGKT 266
 DB 23 -----AVLECTSPIPEARAVLDLDVQCCKQIOGKGFQWVAIEGLDATGKT 68
 QY 267 TVTQSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFYSLGNVIVASEIAKESAKSP 326
 DB 69 TVTQSAADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFYSLGNVIVASEIAKESAKSP 128
 QY 327 VIVDRYWHSTATYATVTEVSGGLOHLPAPHPVYQWPEDLKPDLLILLITVSPPELRQL 386
 DB 129 VIVDRYWHSTATYATVTEVSGGLOHLPAPHPVYQWPEDLKPDLLILLITVSPPELRQL 188
 QY 387 QGRGMEKTRTEAELEANSFRQKVENSYORMENPGCHVVDVADSPSRKVLQTVLSLQNSF 446
 DB 189 QGRGMEKTRTEAELEANSFRQKVENSYORMENPGCHVVDVADSPSRKVLQTVLSLQNSF 248
 QY 447 SEP 449
 DB 249 SEP 251
 RESULT 13
 ID ADF76971
 XX ADF76971 standard; protein; 290 AA.
 AC ADF76971;

XX DT 26-FEB-2004 (first entry)
 XX DE Novel human secreted and transmembrane protein SeqID 646.
 XX KW human; PRO; membrane bound protein; membrane bound receptor;
 KW cell proliferation; cell migration; cell differentiation;
 KW mitogenic factor; survival factor; cytotoxic factor;
 KW differentiation factor; neuropeptide; hormone; cell receptor;
 KW receptor-ligand interaction; cytostatic; chondrocyte; tumour.
 XX OS Homo sapiens.
 XX WO2003072035-A2.
 PD 04-SEP-2003.
 PF 21-FEB-2003; 2003WO-US005241.
 PR 22-FEB-2002; 2002US-0359461P.
 XX (GETH) GENENTECH INC.
 PI Bodary SC, Clark H, Hunte B, Jackman JK, Schoenfeld JR;
 PI Williams PM, Wood WI, Wu TD;
 XX WPI: 2003-721702/68.
 DR N-PSDB; ADF76970.
 XX
 PT New PRO polypeptides, useful for diagnosing and treating an immune
 PT related disorder, e.g. systemic lupus erythematosus, rheumatoid
 PT arthritis, osteoarthritis, juvenile chronic arthritis, thyroiditis or
 XX diabetes mellitus.
 PS Claim 10; SEQ ID NO 646; 918pp; English.
 XX
 CC This invention relates to novel nucleic acids encoding human PRO secreted
 CC and transmembrane proteins. Extracellular proteins play important roles
 CC in the formation, differentiation and maintenance of multicellular
 CC organisms. The fate of many individual cells (for example proliferation,
 CC migration or differentiation) is typically governed by information
 CC received from other cells and the immediate environment. The information
 CC is often transmitted by secreted polypeptides (for example mitogenic
 CC factors, survival factors, cytotoxic factors, differentiation factors,
 CC neuropeptides and hormones) which are received and interpreted by diverse
 CC cell receptors or membrane bound proteins. These membrane bound proteins
 CC as in the blocking of receptor-ligand interactions. The current invention
 CC provides the amino acid sequences of novel human membrane bound receptors
 CC and proteins, along with the cDNA sequences encoding them. The novel
 CC proteins of the invention may have cytostatic activities through the
 CC stimulation of chondrocytes. The nucleic acids of the invention may be
 CC useful for the manufacture of a medicament for diagnosing or treating a
 CC tumour in a mammal. In addition, they may be useful for measuring or
 CC detecting the expression of a tumour associated gene. The present
 CC sequence is the amino acid sequence of a human PRO protein of the
 CC invention.
 XX
 SQ Sequence 290 AA;

Query Match 48.1%; Score 1137; DB 7; Length 290;
 Best Local Similarity 90.9%; Pred. No. 2.1e-106;
 Matches 229; Conservative 3; Mismatches 16; Indels 4; Gaps 2;
 QY 202 PVVPDLPS-SVVPFDPREARAVLECT---SPIPEARAVLDLDVQCCKQIOGKGFQWVAI 257
 DB 39 PIWATLPSWRVSKREAPRRGKSLCTVYLLYIPPEARAVLDLDVQCCKQIOGKGFQWVAI 98
 QY 258 EGLDATGKTTVTQSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFYSLGNVIVASE 317
 DB 99 EGLDATGKTTVTQSVADSLKAVLLKSPSCIGQWRKIFDDEPTIIRAFYSLGNVIVASE 158
 QY 318 IAKESAKSPVIVDRYWHSTATYATVTEVSGGLOHLPAPHPVYQWPEDLKPDLLILLITV 377

Db 159 IAKESAKSPVIVDRYWHSTATYATATEVSGGLQHLPPAHHPVYQWPEDLKPDLLILLTV 218
QY 378 SPEERLQRLQGRMGKMTREAELEANSVFRQKVENSYQRMENPGCHVVDASPSREKVLQT 437
Db 219 SPEERLQRLQGRMGKMTREAELEANSVFRQKVENSYQRMENPGCHVVDASPSREKVLQT 278
QY 438 VLSLQNSFSEP 449
Db 279 VLSLQNSFSEP 290
RESULT 14
ADH42561
ID ADH42561 standard; protein; 200 AA.
XX ADH42561;
AC
XX 25-MAR-2004 (first entry)
DT
DE Novel human protein NOV78b.
XX
XX cardiovascular; antiarteriosclerotic; hypotensive; cytostatic; anorectic;
XX antidiabetic; immunosuppressive; anti-HIV; neuroprotective; nootropic;
XX antiparkinsonian; antischismatic; antiinfertility; cardiomyopathy;
XX atherosclerosis; hypertension; cancer; obesity; diabetes; AIDS;
XX Multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
XX Parkinson's disease; asthma; fertility disorder; chromosome mapping;
XX tissue typing; preventive medicine; pharmacogenomic; vaccine.
OS Homo sapiens.
XX
XX WC2003102159-A2.
XX
XX 11-DEC-2003.
XX
XX 04-JUN-2003; 2003WC-US017573.
XX
XX 04-JUN-2002; 2002US-0385490P.
XX 04-JUN-2002; 2002US-0385615P.
XX 04-JUN-2002; 2002US-0385755P.
XX 05-JUN-2002; 2002US-0386041P.
XX 06-JUN-2002; 2002US-0386355P.
XX 06-JUN-2002; 2002US-0386357P.
XX 06-JUN-2002; 2002US-0386447P.
XX 06-JUN-2002; 2002US-0386459P.
XX 06-JUN-2002; 2002US-0386465P.
XX 06-JUN-2002; 2002US-0386864P.
XX 07-JUN-2002; 2002US-0386701P.
XX 07-JUN-2002; 2002US-0386796P.
XX 07-JUN-2002; 2002US-0386931P.
XX 07-JUN-2002; 2002US-0387078P.
XX 07-JUN-2002; 2002US-0387081P.
XX 07-JUN-2002; 2002US-0387083P.
XX 10-JUN-2002; 2002US-0387429P.
XX 10-JUN-2002; 2002US-0387540P.
XX 10-JUN-2002; 2002US-0387866P.
XX 11-JUN-2002; 2002US-0387610P.
XX 11-JUN-2002; 2002US-0387606P.
XX 11-JUN-2002; 2002US-0387659P.
XX 11-JUN-2002; 2002US-0387668P.
XX 11-JUN-2002; 2002US-0387696P.
XX 11-JUN-2002; 2002US-0387859P.
XX 12-JUN-2002; 2002US-0387934P.
XX 12-JUN-2002; 2002US-0387960P.
XX 12-JUN-2002; 2002US-0388022P.
XX 12-JUN-2002; 2002US-0388096P.
XX 12-JUN-2002; 2002US-0388432P.
XX 12-JUN-2002; 2002US-0388479P.
XX 13-JUN-2002; 2002US-0389123P.
XX 14-JUN-2002; 2002US-0389120P.
XX 14-JUN-2002; 2002US-0389146P.
XX 17-JUN-2002; 2002US-0389742P.

PR 18-JUN-2002; 2002US-0389604P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 25-JUN-2002; 2002US-0391726P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-040268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403458P.
PR 15-AUG-2002; 2002US-0403617P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 23-SEP-2002; 2002US-0412955P.
PR 30-SEP-2002; 2002US-0415195P.
PR 23-OCT-2002; 2002US-0420627P.
PR 23-OCT-2002; 2002US-0420718P.
PR 24-OCT-2002; 2002US-0420852P.
PR 31-OCT-2002; 2002US-0422750P.
PR 01-NOV-2002; 2002US-0423095P.
PR 05-NOV-2002; 2002US-0423748P.
XX (CURA-) CURAGEN CORP.
XX
XX Alsobrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FI;
PI Burgess CE, Casman SU, Caterton E, Dhanabal M, Edinger SR;
PI Ellerman K, Ethenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI Grosse WW, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;
PI Khrantsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;
PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CE;
PI Millet I, Padigaru M, Patturajan M, Peyman JA, Qian X, Rastelli L;
PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;
PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
PI Wolenc AR, Zhong M, Zhong H;
XX
XX WPI; 2004-053467/05.
XX N-PSDB; ADH42560.
XX
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
PT pharmacogenomics.
XX
XX Claim 2; SEQ ID NO 1114; 1503pp; English.
XX
XX The invention relates to 566 new isolated human polypeptides and their
CC encoding genes, sequences that are at least 95% identical to these or
CC sequences comprising one or more conservative substitutions in these. The
CC polypeptide, polynucleotide and antibodies against the polypeptides are
CC useful in diagnosing, treating or preventing NOVX-associated disorders,
CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
CC The nucleic acids are further used as hybridization probes, in chromosome
CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
CC polypeptides are also useful as vaccines. This sequence represents an
CC example of the polypeptide of the invention.
XX
XX Sequence 200 AA;
SQ
Query Match 41.6%; Score 982.5; DB 8; Length 200;
Best Local Similarity 99.0%; Pred. No. 6.1e-91;
Matches 193; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
QY 256 AIEGLDAT-GKTTVTQSVADSLKAVLLKSPSCIGQWRKTFDDEPTIIRAFYSLGNIV 314
Db 3 SIEGLDATGKTTVTQSVADSLKAVLLKSPSCIGQWRKTFDDEPTIIRAFYSLGNIV 62
QY 315 ASEIAKESAKSPVIVDRYWHSTATYATATEVSGGLQHLPPAHHPVYQWPEDLKPDLL 374
Db 63 ASEIAKESAKSPVIVDRYWHSTATYATATEVSGGLQHLPPAHHPVYQWPEDLKPDLL 122

Db 123 LTVSPEERLQRLQGRGMEKTRREEALEANSVFRQKVENSYQRMENPGCHVVDASPSREKV 182
Qy 435 LQTVLSLIONSFSEP 449
Db 183 LQTVLSLIONSFSEP 197

Search completed: October 22, 2004, 23:30:58
Job time : 123 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 22, 2004, 10:45:02 ; Search time 7916 Seconds

(without alignments)
11835.087 Million cell updates/sec

Title: US-10-681-223-1

Perfect score: 2571

Sequence: 1 cggggcggaggcgctcg.....aggaggggcttttcacccc 2571

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gssi:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	885.2	34.4	1003	5	BX463706	BX463706 BX463706
2	884.2	34.4	324	3	BC027329	BC027329 Mus muscu
3	873.2	34.0	3214	3	BC048910	BC048910 Mus muscu
4	872.4	33.9	3022	3	AK004595	AK004595 Mus muscu
5	771.4	30.0	1084	5	BX421924	BX421924 BX421924
6	771.4	30.0	1287	3	CR623316	CR623316 full-leng
7	741.8	28.9	949	6	CF272319	CF272319 AGENCOURT
8	690.2	26.8	757	4	BG501037	BG501037 602546443
9	685.2	26.7	767	4	BG497576	BG497576 601859469
10	674.8	26.2	759	4	BG497578	BG497578 602543017
11	653.2	25.4	769	5	BUI83182	BUI83182 AGENCOURT
12	646.6	25.1	705	4	BG778349	BG778349 602666287
13	646.6	25.1	833	4	B1551715	B1551715 603197411
14	621.8	24.2	844	4	BG778145	BG778145 602666482
15	613.8	23.9	720	4	BG496231	BG496231 602538182
16	607	23.6	1262	5	BQ067262	BQ067262 AGENCOURT
17	605.4	23.5	801	5	BUI60713	BUI60713 AGENCOURT
18	602.4	23.4	611	5	BU070188	BU070188 in19c07.y
19	600.4	23.4	972	4	BUI78500	BUI78500 602328334
20	594.4	23.1	596	6	CB114648	CB114648 K-EST0158
21	588.4	22.9	868	2	BF240527	BF240527 601875619
22	588.2	22.9	669	2	BE748448	BE748448 601571914
23	584.2	22.7	1333	4	BM906075	BM906075 AGENCOURT
24	583.6	22.7	765	2	BF241358	BF241358 601876319

25	569	22.1	728	4	BG779993	BG779993 602668235
26	563.6	21.9	751	2	BF102537	BF102537 601646494
27	563.6	21.9	868	2	BF243467	BF243467 601878183
28	562	21.9	700	4	B1818166	B1818166 603032614
29	553.6	21.5	615	4	BG776607	BG776607 602663751
30	549.4	21.4	616	1	AV707941	AV707941 AV707941
31	549.2	21.4	813	2	BF242831	BF242831 601878235
32	534	20.8	702	4	BG776636	BG776636 602663790
33	533.2	20.7	795	4	BG771320	BG771320 602719573
34	531.8	20.7	595	4	BG501398	BG501398 602547746
35	529.8	20.6	837	2	BF242245	BF242245 601878076
36	524.2	20.4	669	4	BG497460	BG497460 601859520
37	521.8	20.3	704	4	BG547557	BG547557 602575410
38	520.2	20.2	1016	6	BY704814	BY704814 BX704814
39	516.6	20.1	596	2	BE748211	BE748211 601571424
40	515.8	20.1	745	2	BF031764	BF031764 601558088
41	513.8	20.0	830	5	BUI66495	BUI66495 AGENCOURT
42	511.2	19.9	735	2	BF243749	BF243749 601877866
43	506	19.7	1008	4	B1918806	B1918806 603180784
44	493.8	19.2	928	4	BG180852	BG180852 602327301
45	483.8	18.8	771	4	BG255761	BG255761 602368237

ALIGNMENTS

RESULT 1
BX463706
LOCUS
DEFINITION
ACCESSION
VERSION
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BX463706 Homo sapiens ADULT BRAIN Homo sapiens linear EST 06-MAY-2004
CS0DN005YL22 5-PRIME, mRNA sequence.
BX463706
BX463706.2 GI:47071037
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 22, 2003 this sequence version replaced gi:31035556.
Contract: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 1573.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DN005YF11Q1P1sc=1573.r.

FEATURES

source
1..1003
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DN005YL22"
/tissue_type="ADULT BRAIN"
/clone_stage="adult"
/clone_lib="Homo sapiens ADULT BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN

Query Match 34.4%; Score 885.2; DB 5; Length 1003;
Best Local Similarity 99.0%; Pred. No. 3.2e-198;

Matches	901;	Conservative	0;	Mismatches	8;	Indels	1;	Gaps	1;
Qy	695	GTCTGGAGAGTGTACCTCTTTATCTCTGAAGCCCGGCGAGTGTCTGACCTGTGTCGAC	754						
Db	95	GCTGTGACAGTGTACCTCTTTATCTCTGAAGCCCGGCGAGTGTCTGACCTGTGTCGAC	154						
Qy	755	CAGTGCCTCCAAACAGATCCAGAAAGGAATGTCAGGTTGTCCTATCGAAGGACTGGAT	814						
Db	155	CAGTGCCTCCAAACAGATCCAGAAAGGAATGTCAGGTTGTCCTATCGAAGGACTGGAT	214						
Qy	815	GCACGGGTAAACACAGCTGACCCAGTCAGTGGCAGATTCACCTTAAGGCTGTCTCTCTTA	874						
Db	215	GCACGGGTAAACACAGCTGACCCAGTCAGTGGCAGATTCACCTTAAGGCTGTCTCTCTTA	274						
Qy	875	AAGTCACCCCTCTTTCATTTGGCCAGTGGAGGAAGATCTTTGATGATGAACCAACTATC	934						
Db	275	AAGTCACCCCTCTTTCATTTGGCCAGTGGAGGAAGATCTTTGATGATGAACCAACTATC	334						
Qy	935	ATTAGAAGAGCTTTTACTCTTTGGGCAATTATATTGTGGCCCTCCGAAATAGCTAAGAA	994						
Db	335	ATTAGAAGAGCTTTTACTCTTTGGGCAATTATATTGTGGCCCTCCGAAATAGCTAAGAA	394						
Qy	995	TCGTCCAAATCTCTGTGATTGTAGACAGTACTGGCAGACGCGCCACTATGCCATA	1054						
Db	395	TCGTCCAAATCTCTGTGATTGTAGACAGTACTGGCAGACGCGCCACTATGCCATA	454						
Qy	1055	GCCACTGAGTGTGGGGTCTCCAGACCTGCCGCCAGCCCATCCCTGTGTACCAG	1114						
Db	455	GCCACTGAGTGTGGGGTCTCCAGACCTGCCGCCAGCCCATCCCTGTGTACCAG	514						
Qy	1115	TGCCCAGAGACTGCTCAAACTGACCTTATCTGTCTGCTCACTGTGAGTCTGAGGAG	1174						
Db	515	TGCCCAGAGACTGCTCAAACTGACCTTATCTGTCTGCTCACTGTGAGTCTGAGGAG	574						
Qy	1175	AGTTTGAGAGGCTGACGGCCGGGCGATGAGAAAGACACGAGGAAGAGCAACTGTGAG	1234						
Db	575	AGTTTGAGAGGCTGACGGCCGGGCGATGAGAAAGACACGAGGAAGAGCAACTGTGAG	634						
Qy	1235	GCCAAAGTGTGTTTGTGTCAAAAGGTAGAAAATGTCTTACAGCGGATGGAGAAATCTCTGGC	1294						
Db	635	GCCAAAGTGTGTTTGTGTCAAAAGGTAGAAAATGTCTTACAGCGGATGGAGAAATCTCTGGC	694						
Qy	1295	TGCCATGTGTTGTTTGTGTCAAAAGGTAGAAAATGTCTTACAGCGGATGGAGAAATCTCTGGC	1354						
Db	695	TGCCATGTGTTGTTTGTGTCAAAAGGTAGAAAATGTCTTACAGCGGATGGAGAAATCTCTGGC	754						
Qy	1355	ATCCAGAAATAGTTTGTAGTGAACCGTAGTTACTCTGGCCAGGTGCCAGTCTAACTAGATT	1414						
Db	755	ATCCAGAAATAGTTTGTAGTGAACCGTAGTTACTCTGGCCAGGTGCCAGTCTAACTAGATT	813						
Qy	1415	AGATGTTGTTTGAACATCTACATCCACCATTTGTTATGAGTGTTCGAAATTTCTGTT	1474						
Db	814	AGATGTTGTTTGAACATCTACATCCACCATTTGTTATGAGTGTTCGAAATTTCTGTT	873						
Qy	1475	CTACAAGCATGTTGTGGCAGAAAACCTGGAGACAGGCATCTTAATTTTACTTCAGCCA	1534						
Db	874	CTACAAGCATGTTGTGGCAGAAAACCTGGAGACAGGCATCTTAATTTTACTTCAGCCA	933						
Qy	1535	TCGTACCTCTTCTGATGTGAGCCCGTCAATCAAAAGGTGCCCTCTCATCATGTTCCAG	1594						
Db	934	TCGTACCTCTTCTGATGTGAGCCCGTCAATCAAAAGGTGCCCTCTCATCATGTTCCAG	993						
Qy	1595	TGAGAGGCCA 1604							
Db	994	TGAGAGGCCA 1003							

RESULT 2
BC027329
LOCUS
DEFINITION
ACCESSION

BC027329
Mus musculus, similar to thymidylate kinase family LFS-inducible member, clone IMAGE:3982036, mRNA.

3124 bp
mRNA
linear
HTC 20-SEP-2002

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

BC027329.1 GI:20072707
HTC
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3124)
Strausberg, R.
Direct Submission
Submitted (04-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: angbcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

REMARK
COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 36 Row: e Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein
This clone has the following problem: frame shifted.

FEATURES
source

Location/Qualifiers
1..3124
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:3982036"
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"

ORIGIN

Query Match 34.4%; Score 884.2; DB 3; Length 3124;
Best Local Similarity 77.0%; Pred. No. 7.2e-198;
Matches 1121; Conservative 0; Mismatches 318; Indels 16; Gaps 3;
Qy 46 CCGGCTCTCGGGGGCCACTGTGGGGCCGCTCTCGGGCGGCGGGGCTCGGGCTGG 105
Db 63 CCCACTGTGGCTGGGGCGTCTGTGCGGAGGCTGTGGGGCGACACAGGCTCGATGGG 122
Qy 106 GGCATGGCTCCGGCGTGGCGCTTCCTGCTGGAGCTTCCGACTGCACTCGGCTCACT 165
Db 123 GACCATGGCTCGGGCGGCGCTTCACTGTGGAGCTGCCAGATTGCTCCCTGACTCACT 182
Qy 166 CGCCCTAGGCGCGACGCCCGCGGCGAGAGCGCCCGCCCGCTCGGGGGCGCT 225
Db 183 CGTCTGGG-----GGATGCAACAGACACAGGATGCACGCTGGCAGAGCT 230
Qy 226 GTTGGGGCCCCCGAGGCGCACTACTCGTGTGGTGGTCCCGTACCCCGAGCGGGCTG 285
Db 231 GCTCGGGCCCCCGAGGCGCACTACGCGCTGTGTGTGCCCCCTGGCTCCAGGCGAAGCTG 290
Qy 286 CGGGGCGGGTCCGGGGCGGCGCTGACACAGCGCTGTGACACAGCTCGCGCGGGG 345
Db 291 CGGGCCCCGGGTGAGGCGGGCGGGGTGCACATCGCTGTGTAGCAGCTCGCGCGGGG 350


```
QY 346 CCCCTTCACGCGGTGCACAGCTGCTAGGCTGCTGTCTGCTACTGCCCCGGGGCGGCAGCGCCGG 405
Db 351 TCCCTTACAGAGATCCAACTTCTGAGCAAGCTTCTGGGCTATGTGCTCCGGGCGATCAGGCCGG 410
QY 406 CGGCGCACAGCAGGCTTCTGCTGGCGACCCCTGGATGACCTCGACACCCGCGCAAGC 465
Db 411 CCAAGCCACAGCATGGCTTCTGCTGGGACCTTTCGACCCCGGACACTCGGCGCA 470
QY 466 GCTGCTCGAGCTGCTGGGCGCTTCCAGGAGGACCAACCCCGCACTTGGGCGAGTTGCA 525
Db 471 CTTGCTCCAGCTTCTGGGCTCTCTGCCAGAGCGCGCGCCGCACTTGGCGAGTTCCA 530
QY 526 GCGCGACCCGCGGCGACAGCTGGAGCGCTTGGAGAGTGCAAGCGGCGGCT 585
Db 531 GCGCGACTCCCGAGGTTTCTGCTGGCAGCGTCTGTGGAGTTGCAAGGAGCAGGCGAGT 590
QY 586 GCAAGTGGGCTGCGCACAGTGTGCGCGCTCCGGAGCCCGCTGCACCCGCTGTGCG 645
Db 591 GCAAGTGGAGTGGCGATGCTGCTGCCGCAAGGAACTCTATCTGCACCCATTTGCTGCC 650
QY 646 AGACTTGCACAGTTCGCTGCTTCCCGACCGGGAAGCGCGCGCGCTTTTGGAGGA 705
Db 651 AGATCTGCTTAATCTGCGGTGTTCAGACCGGAGCGCGCAAGGCTGTCTTGGAGA 710
QY 706 GTGTACTCTCTTATTCCTGAAGCCCGGCGAGTCTTGACTGGTGCAAGCTGCCCCAA 765
Db 711 GTGCATCTCTTATTCAGAGACCGCGGCACTTGTGACCTAGTTGACAGTGCCTCAA 770
QY 766 ACAGATCCAGAAAGAAAGTTTCAGGTTCTTGCCATCGAAGGACTGGATGCCAGGGTAA 825
Db 771 GGAGGTCCAGAAAGGAAAGTTCCAGTCAITGCCATTTGAAGACTGGATGCCAGTGA 830
QY 826 AACACGGTGAACCGAGTCAAGTGCAGATCACTTAAGGCTGCTCTTAAAGTCAACACC 885
Db 831 GACCACACTGACGCACTGAGTGCAGTCTCTCAAGGCTGCTCTCTACAGTGCACCC 890
QY 886 CTCCTGCAATGCGCAGTGGAGGAGTCTTGATGATGAACCAACTATCATTTAGAAGAG 945
Db 891 CCGCTGTATCAGCACTGGAGGAAGATCTTTGATGATGAACCACTATCATTTAGAAGAG 950
QY 946 TTTTACTCTTTGGGCAATTTATTTGTCGCTCCGAAATAGCTTAAAGATCTGCCAAATC 1005
Db 951 CTTTATTTCTTTGGGCAATTTCTGCGTCTGAAATAGCTTAAAGATCAACCAACTT 1010
QY 1006 TCTGTGATTTAGACAGTACTGGCACAGCGCCACCTATGTCATAGCCACTGAGT 1065
Db 1011 TCTGTGTTATTTAGACAGTACTGGCATAGCACAGCCACCTAGCCCATAGTACTGAGT 1070
QY 1066 GAGTGGGCTCTCCAGACCTGCCCCCGCCATCACCCCTGTGTACCTAGTGGCCAGAGA 1125
Db 1071 GAGTGGAGCCCTACAGTACTACCCCTGCCACCACTGTGTACCTAGTGGCCAGGGA 1130
QY 1126 CTTGCTCAAACTGACTTATCTGCTGCTCACTGTGAGTCTCTGAGGAGAGTTGACAG 1185
Db 1131 CTTGCTGAAGCCGACTTGGTCTGCTGCTGACTGTGAATCTTGAGGAGAGTGGCGAG 1190
QY 1186 GTTGAGGCGCGGGCAUUGAGAACACAGGGAAGAGCAGAACTTTGAGGCCAACAGTGT 1245
Db 1191 ACTGAGGCGCGGGTCAGGAGAAAACTAAAGAAAGAGGTGAACCTTTGAGGCCAAATATGT 1250
QY 1246 GTTTCGTAAGAGGTAGAATGCTTACAGCGGATGGAGATCTCTGGCTGCCATGTGT 1305
Db 1251 GTTTCGTCAGAAGGTGGAATGACGTACACGCGTATGGAGAACCCCAAGCTGCCATCTGT 1310
QY 1306 TGATGCAGCCCTCCAGAGAAAGGTCTCTGAGACAGTATTAAGCCTTAATCCAGATAG 1365
Db 1311 GATGCGACGCCCTCGAGAGAGACAGTCTCTGAGAGAGTTTATGAGCTGATCCAGAGTTC 1370
QY 1366 TTTTGTAGAACCGTAGTTACTCTGGCCAGGTGTCACGCTCTAACTAGATTAGTTGTTT 1425
Db 1371 TGTGCTGTTAAATGTAGTTGTTCTGGCCATGTTCTCTCATTTACACCGG---AGAAGCTGTGT 1427
```

```
QY 1426 GAAACATCTACATCCACCATTTGTTATGCAAGTGTTCGCCAAATTTCTGTTCTTACAGCATG 1485
Db 1428 GATGCCCGGAAACCCATCATCTTCCATGCGAG-CITCTTAGATTATGGGGGTATGAGCATA 1486
QY 1486 TTGTGTGGCGAGAAA 1500
Db 1487 TTGTGTGGNAAAAA 1501

RESULT 3
BC048910
LOCUS Mus musculus, clone IMAGE:6494087, mRNA.
DEFINITION BC048910
ACCESSION BC048910.1 GI:29294752
VERSION HTC.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3214)
Direct Submission
Submitted (17-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kerteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Place: 105 Row: p Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein
This clone has the following problem: frame shifted.

FEATURES
Location/Qualifiers
source
1..3214
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6494087"
/tissue_type="Eye, retina, mouse strain C57Bl/6"
/clone_lib="NIH MGC_94"
/lab_host="DH10B"
/vector="Vector: PCMV-SPORT6"

ORIGIN
Query Match 34.0%; Score 873.2; DB 3; Length 3214;
Best Local Similarity 77.0%; Pred. No. 2.9e-195;
Matches 1121; Conservative 0; Mismatches 318; Indels 17; Gaps 4;

QY 46 CCGGCTCTGCGCGGCCACTGTCTGGGCGCGCTGTCTGGGCGCGCGGGTCTGCGCTGG 105
Db 175 CCCACTGTGGCTCGCGCGCTGTCTGCGGAGGCTGTGCGGCGCACACGAGCCTGCATGG 234
QY 106 GCGCATGGCTCGCGCTGCGCGCTTCTGCTCGAGCTTCCGACTGCACTCGCTGCTCACT 165
Db 235 GACCATGGCTCGCGCGCGCGCTTCACTGTGAGCTGCAGATTGCTCCCTGACTCACTT 294
QY 166 CCGCCTAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 225
Db 295 CGTCTCGG-----GGATGCAACACACAGGATGACCGCTGGCAGAGCT 342
```


Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resesgsc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5', GAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGGATCCAAAGCTCAATTAATTAATTAACCCCGCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

source

Location/Qualifiers
1..3022
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:1200004E04"
/db_xref="taxon:10090"
/clone="1200004E04"
/sex="male"
/tissue type="lung"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
129..1316

CDS

/note="unnamed protein product; putative
thymidylate kinase family LPS-inducible member
(MGDI:99830)"
/codon_start=1
/protein_id="BAB23396.1"
/db_xref="GI:12835871"
/translation="MQQTGMHAWSCSAPQGAATRLVPLAPGEGCGPRVQAAVHH
RLUQLRRFLQRCQLSKLLGYPGDQAGEAQLFLRDPDHPDRLDQLLGLSCQ
EAARPLAEFQADSGLLMQRLEWLGQDQVDCVLPAPQEPHLPLDLLNSAV
FQDRAARAVLEECTFIPEARVLDLVQCPKQVKGQFQVIAEGLDQTKTLTQ
SVSESLKAVLLSPPCISOWRKI FDEPTIIRAFYSIGNYLVAIEAKESTNPVI
VDYKHSTATYAIANEVSGLOVLPAPHVQWPCDILKPLVLLITVNSERVRL
QGRGEKTEAELEANNVFRKVENTYORMENPSCHLVDSFRETVLQKVLLELIQS
SGR"
3000..3005
/note="putative"
3022
/note="putative"

polyA_signal

polyA_site

ORIGIN

Query Match 33.9%; Score 872.4; DB 3; Length 3022;
Best Local Similarity 77.1%; Pred. No. 4.4e-195;
Matches 1119; Conservative 0; Mismatches 316; Indels 17; Gaps 4;
QY 50 CTCCTCGGGGGCCACTGTCGGGGCGCGTGTCTGGCGGGCGGGGTCTGCGCTGGGGCC 109
DB 2 CTGTGGCTGGGGGTCTGTGCGGAGGCTGTGCGGGCGACACAGGCGCTGCATGGGGACC 61
QY 110 ATGGCTCCCGCGCGCTCGTCCCTGGAGCTCCCGACTGCACCCCTGGCTCACTTCGCC 169

DB 62 ATGGCTCGCGCGGGGGCTTCACTGTGGAGTCGCAGATTGCTCCCTGACTCACTTCGTC 121
QY 170 CTAGGCGCGACGCGCCCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 229
DB 122 CTGGG-----GGATCAACAGACACACAGGATGCAACGCGCTGGCAGAGCTGCTC 169
QY 230 GGGCCCCCGAGGAGGAGCTACTCG-CTGTGCGTGGCCGTGACCCCGGACGCGCGCTGGCG 288
DB 170 GGGCCCCCAGGCGGAGCTACGCGCTGTGTGCGCCCTGGCTCCAGGCGAAGGCTGGCG 229
QY 289 GGGCGGGTCCCGGGCGCGCGCTGCACAGCGCTGTGTCACACAGCTCGCGCGCGCGCC 348
DB 230 GGGCGGGTGCAGGCGGCGCGGTGCACATCGCTGTGTCAGCAGCTGCGCGGGGTCC 289
QY 349 CTTCCAGCGGTGCCAGCTGCTCAGGCTGCTGTGCTACTGTGCGCGGGCGGCGACGCGGGCG 408
DB 290 TTTACAGAGATGCCAACTGAGCAAGCTTCTGGGCTATGCTCGGCGCATCAGGCGCGCGCA 349
QY 409 CGCAGCAGNAGCTTCTGCTGGGCGACCCCTGTGATGACCTGCACCCGCGCAAGCGCT 468
DB 350 AGCCGAGCATGCTTCTGCTGCGGACCCCTTGGCACACCCGCGACACTCGCGCGGACTT 409
QY 469 GCTCAGCTGCTGGGCGCTGCGAGGACACACGCGCGCACTTGGGCGAGTTCGAGGC 528
DB 410 GCTCAGCTTCTGGGCTCTGCGAGGCGGCGCGCGCTGCTGGCGAGTTCAGGC 469
QY 529 CGACCGCGCGCGCAGCTGTGCGAGCGCTCTGGAGGTCGCAAGCGGCGAGCGGCTGCA 588
DB 470 CGACTCCAGGCTTGTGCTGGCAGCGCTGTGGGAGTTGCGAGGAGACAGCGAGGTGCA 529
QY 589 GGTGGCTGCGCACAGGTGCTGCGCGGCGCGCTGCGAGCGCGCGCTGCACCCGCTGGTGC 648
DB 530 GGTGACTCGCATGCTGCTGCGGCGACAGGAACTCATCTGCACTTCGTCGCGA 589
QY 649 CTTGCCAGTTCGCTGGTCTTCCGACCGGAAAGCGCGCGCGCTTTGGAGAGTG 708
DB 590 TCTGCTTAATCTGCGGTGTTCCAGACCGGACGCGGCAAGGCTGTCTTGAAGAGTG 649
QY 709 TACCTCCTTTATTCGAGCGCGGCGAGTGTGACCTGTGACCTGTGACCTGTGACCTGTG 768
DB 650 CACATCCTTTATTCAGAAAGCGCGGCGAGTCTTGAAGCTGTGACCTGTGACCTGTGAC 709
QY 769 GATCAGAAAGAAAGTTCAGGTTGTTGCCATCGAAGACTGGATGCCACGGTAAAC 828
DB 710 GGTCCAGAAAGGAAAGTTCAGGTTTCCAGGTCATTGCCATTTGAAGACTGGATGCCAT 769
QY 829 CACGCTGACCCAGTCAAGTGGCAGATTCACTTAAGGCTGTCTTAAAGTCAACACCTC 888
DB 770 CACACTGACGAGTCAAGTGTGAGAGTCTCTCAAGGCTGTCTCTACAGTCCGACCC 829
QY 889 TTGCATTTGCCAGTGGAGGAAGATCTTGATGATGAACCAATATCATTTAGAGAGCTTT 948
DB 830 CTGTATCAGCCAGTGGAGGAAGATCTTTGATGATGAACCTACTATCATTCGAAGAGCTT 889
QY 949 TTACTCTTTGGGCAATATATTGTGGCTCCGAAATAGCTAAAGATCTGCCAATCTCC 1008
DB 890 TTATTCTTTGGGCAATATCTCGTGGCTTCTGAAATAGCTAAAGATCAACCACTTCC 949
QY 1009 TGTGATTTGACACAGGTACTGCGACAGCAGCGCCACTATGCGCATAGGCACTGAGGTGAG 1068
DB 950 TGTATTGTAGACAGGTACTGCGCATAGCACGCCACCTAGCGCATAGCTACTGAGGTGAG 1009
QY 1069 TGGGGGTCTCCAGCACCTGCGCCCGAGCCCATCACCTGTGTACCTAGTGGCCAGAGACCT 1128
DB 1010 TGGAGGCTTACAGTACCTACCCCTGCCACCCCTGTGTACCTAGTGGCCAGGAGCT 1069
QY 1129 GCTCAAACTGACCTTATCTGCTGCTCACTGTGAGTCTGAGGAGAGGTTTCAGAGGCT 1188
DB 1070 GCTGAAGCCGACTTGGTCTGCTGTGAATCTGAGGAGAGAGTTCGGAGCT 1129
QY 1189 GCAGGCGGGGCATGGAGAACACAGGGAAGACAGAACTTGAGGCCAACAGTGTGT 1248
DB 1130 GCAGGCGGGGTGAGGAGAAACTTAAGAGAGGCTGAACCTTGAGGCCAATATGTGT 1189

QY 1249 TCGTCAAAAGGTAGAAATGCTCTACAGCGGATGGAGATFCTGGCTGCCATCTGGTGA 1308
 |||||
 Db 1190 TCGTCAGAAAGGTGGAATGACGTACCGGTATGGAGAACCCCAAGCTGCCATCTGGTGA 1249
 |||||
 QY 1309 TCCAGCCCTCCAGAGAAAGGTCTCGCAGACAGTATTAAAGCTAATCCAGATAGTTT 1368
 |||||
 Db 1250 TGCAGCCCTCCAGAGACAGTCTCGCAGAGAGGTTTATAGAGCTGATCCAGAGTTCTGG 1309
 |||||
 QY 1369 TAGTGAACCGTAGTACTCTGGCCAGGTCACAGTCTAACTAGATTAGATTGTTTGA 1428
 |||||
 Db 1310 TCGTTAATTGTAGTTCTGGCCATGTTCTCAATTACACCGG--AGAAGCTGTGTGAT 1366
 |||||
 QY 1429 ACATCTACATCCACCAATTTGTTAAGCAGTGTCCCAAAATTTCTGTCTCAACAGCATGTTG 1488
 |||||
 Db 1367 GCCCCGAAACCCATCACTTGCCATGCAG-CTTCTTAGATTAGGGGTATGACCATATTT 1425
 |||||
 QY 1489 TGTGGCAGAAAA 1500
 |||||
 Db 1426 GGTGGAATAAAA 1437
 |||||

RESULT 5
 BX421924
 LOCUS
 DEFINITION BX421924 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cDNA
 clone CS0DH005YN12 5-PRIME, mRNA sequence.
 ACCESSION BX421924
 VERSION BX421924.2 GI:46929797
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1084)
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 13, 2003 this sequence version replaced gi:30647154.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
 was not normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 This sequence belongs to sequence cluster 1572.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CS0DH005DG06QPl&c=1572.r.

FEATURES
 source
 1..1084
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DH005YN12"
 /cell_line="JURKAT CELL LINE"
 /tissue_type="T CELLS (JURKAT CELL LINE)"
 /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"
 /notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN
 Query Match 30.0%; Score 771.4; DB 5; Length 1084;
 Best Local Similarity 98.5%; Pred. No. 2.8e-171;
 Matches 852; Conservative 0; Mismatches 6; Indels 7; Gaps 7;
 QY 1629 TAAACATTTCTTGGACATATGTTTCACCTTAATCACTACCAATATCTGGAAGACCTGT 1688
 |||||

Db 1 TAAACATTTCTTGGACATATGTTTCACCTTAATCACTACCAATATCTGGAAGACCTGT 60
 QY 1689 CTTACTCAGACAGCACCAGTGTACAGAGACGACAGCAAGATCTTCCAGATCAGCAGG 1748
 |||||
 Db 61 CTTACTCAGACAGCACCAGTGTACAGAGACGACAGCAAGATCTTCCAGATCAGCAGG 120
 |||||
 QY 1749 AGACCCCGAGAGCTCTGCTTCTCTACACTGGCATCTCATGAGATCGTGACATGCCAC 1808
 |||||
 Db 121 AGACCCCGAGAGCTCTGCTTCTCTACACTGGCATCTCATGAGATCGTGACATGCCAC 180
 |||||
 QY 1809 ATTGGCTTCTTCCACATCTGGTTGCACTCGTCATGATGGCTGCGTGCATCTCCCTCAGT 1868
 |||||
 Db 181 ATTGGCTTCTTCCACATCTGGTTGCACTCGTCATGATGGCTGCGTGCATCTCCCTCAGT 240
 |||||
 QY 1869 CCCAAATCTTAGTAGCAGAGTGTCTCTGAGAGGCTCTTATGTGTCCTGGCTGCCAAG 1928
 |||||
 Db 241 CCCAAATCTTAG-AGCCAAAGTGTCTCTGAGAGGCTGTCTATGTGTCCTGGCTGCCAA- 298
 |||||
 QY 1929 GGACACTCTCTGAGAGGCCATTTTGGGTAAGGAACAACCTTACAAAGAGGCATTGATCTTG 1988
 |||||
 Db 299 GGACACTCTCTGAGAGGCCATTTTGGGTAAGGAACAACCTTACAAAGAGGCATTGATCTTG 358
 |||||
 QY 1989 TGTCTGAGGCTCAGAGCCCTTTTGTAGTAGGCTTCTGATGTCAATTCATAAAGACATTCAGC 2048
 |||||
 Db 359 TGTCTGAGGCTCAGAGCCCTTTTGTAGTAGGCTTCTGA-GTCATATATAAAGACATTCAGC 417
 |||||
 QY 2049 CAAGATGCTCCAACTGCAATATACCAACCTTCTCTGAAATATATTTTGTCTTATTTATAT 2108
 |||||
 Db 418 CAAGATGCTCCAACTGCAATATACCAACCTTCTCTGAAATATATTTTGTCTTATTTATAT 477
 |||||
 QY 2109 TTCTTTTCTTTTCTTAAAGAAATGGCTCTGAATAGAAATGCACATTTTCCATCTGAAC 2168
 |||||
 Db 478 TTCTTTTCTTTTCTTAAAGTA-TGGCTCTGAATAGAAATGCACATTTTCCAT-TGAAC 535
 |||||
 QY 2169 GATGATATCATTTAGCCAAATCCAGTAATTTATTTATTTATTAATCTATACATATATGTT 2228
 |||||
 Db 536 GGATGATTTCAATTTAGCCAAATCCAGTAATTTATTTATTTATTAATCTATACATATATGTT 595
 |||||
 QY 2229 TCCTCAGCATAGAGCTATGATTCATTTAAAGTGGAGTCAAAACGCTAAATGCAAT 2288
 |||||
 Db 596 TCCTCAGCATAGAGCTATGATTCATTTAAAGTGGAGTCAAAACGCTAAATGCAAT 655
 |||||
 QY 2289 GTTTGTTGTATTTTCAATACACAACTTAATTTGTTTGTAAATAGTCAAGTGA 2348
 |||||
 Db 656 GTTTGTTGTATTTTCAATACACAACTTAATTTGTTTGTAAATAGTAC-AGTGA 714
 |||||
 QY 2349 TCTTGAGTGGGATTTCTTGGTAAATTAATCTTGCACTTGAATGTCTCATGATACATATG 2408
 |||||
 Db 715 TCTTGAGTGGGATTTCTTGGTAAATTAATCTTGCACTTGAATGTCTCATGATACATATG 774
 |||||
 QY 2409 AATTCGTTTGACATATCTTTAGACAGAAAAAGTAGCTGAGTGGGGGAAATATATAGA 2468
 |||||
 Db 775 AATTCGTTTGACATATCTTTAGACAGAAAAAGTAGCTGAGTGGGGGAAATATATAGA 834
 |||||
 QY 2469 GCTTGTGTGACTTTAGGGAGTAGCT 2493
 |||||
 Db 835 GC-TGTTGACTTTAGGGAGTAGGT 858
 |||||

RESULT 6
 CR623316
 LOCUS
 DEFINITION full-length cDNA clone CS0DH005YN12 of T cells (Jurkat cell line)
 of Homo sapiens (human).
 ACCESSION CR623316
 VERSION CR623316.1 GI:50504123
 KEYWORDS HTC; CNSLT_cDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1287)
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 CR623316 1287 bp mRNA linear HTC 21-JUL-2004
 full-length cDNA clone CS0DH005YN12 of T cells (Jurkat cell line)
 of Homo sapiens (human).
 CR623316
 CR623316.1 GI:50504123
 HTC; CNSLT_cDNA.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1287)
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

```

TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished
REMARK     Contact : Feng Liang Email : fliang@lifetech.com URL :
           http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
           Faraday Avenue
REFERENCE  2 (bases 1 to 1287)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage :
           BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
           - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
           end enriched, double-strand cDNA was digested with Not I and cloned
           into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
           was normalized. Library was constructed by Life Technologies, a
           division of Invitrogen.
FEATURES   Location/Qualifiers
           1..1287
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="CS0DH005YN12"
           /tissue_type="T cells (Jurkat cell line)"
           /plasmid="pCMVSPORT_6"

ORIGIN
Query Match      30.0%; Score 771.4; DB 3; Length 1287;
Best Local Similarity 98.5%; Pred. No. 2.9e-171;
Matches 852; Conservative 0; Mismatches 6; Indels 7; Gaps 7;

QY 1629 TAAACATTTCTTGAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGT 1688
Db 1 TAAACATTTCTTGAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGT 60

QY 1689 CTTACTCAGACAGCACCAGGTGTACAGAAGCAGCAGCAAGATCTTCCAGATCAGCAGGG 1748
Db 61 CTTACTCAGACAGCACCAGGTGTACAGAAGCAGCAGCAAGATCTTCCAGATCAGCAGGG 120

QY 1749 AGACCCGGAGGCTCTGCTTCTCCTACATGTCATGTCATGATGGCTCGCTGCTCCCTCAGT 1808
Db 121 AGACCCGGAGGCTCTGCTTCTCCTACATGTCATGTCATGATGGCTCGCTGCTCCCTCAGT 180

QY 1809 ATTGGCTTCTCCACATCTGTTTGACTCGTCTGATGATGGCTCGCTGCTCCCTCAGT 1868
Db 181 ATTGGCTTCTCCACATCTGTTTGACTCGTCTGATGATGGCTCGCTGCTCCCTCAGT 240

QY 1869 CCCAAATCTAGTAGCAGAGTGTCTCTGAGAGGCTGTCTATGTCTCTGGCTGCCCAAG 1928
Db 241 CCCAAATCTAG-AGCCAAAGTGTCTCTGAGAGGCTGTCTATGTCTCTGGCTGCCCA- 298

QY 1929 GGACACTCTGAGAGGCCATTTTGGTAAGGAACACTTACAAAGAGGCATTTGATCTTTG 1988
Db 299 GGACACTCTGAGAGGCCATTTTGGTAAGGAACACTTACAAAGAGGCATTTGATCTTTG 358

QY 1989 TGTCTGAGGCTCAGAGCCCTTTTGATAGGCTTCTGATGTCTTATTAAGACATTCACGC 2048
Db 359 TGTCTGAGGCTCAGAGCCCTTTTGATAGGCTTCTGATGTCTTATTAAGACATTCACGC 417

QY 2049 CAAGATGCTCCAACTGCAATATACCAACCTTCTCTGAATATATTTTCTTATTTATAT 2108
Db 418 CAAGATGCTCCAACTGCAATATACCAACCTTCTCTGAATATATTTTCTTATTTATAT 477

QY 2109 TTCTTTTCTTTTCTTAAAGAAATTTGGCTCTGATGAATGACATTTTCCATCTGAAT 2168
Db 478 TTCTTTTCTTTTCTTAAAGTA-TGGCTCTGAATGAATGACATTTTCCAT-TGAAT 535

QY 2169 GGATGCATATCATTTAGCCAACTCCAGTATTTTATTTATTTATTAATCATACATAATGTT 2228
Db 536 GGATGCATTTCAATTTAGCCAACTCCAGTATTTTATTTATTTATTAATCATACATAATGTT 595

QY 2229 TCCTCAGCATAGGAGCTATGATTCATTAATTAAGCTGGAGTCAAAACGCTAAATGCAAT 2288
Db 596 TCCTCAGCATAGGAGCTATGATTCATTAATTAAGCTGGAGTCAAAACGCTAAATGCAAT 655

```

```

QY 2289 GTTCTGTGTTATTTTCATTACACAACTTAATTTGTCTTGTAAATAGTTCAAGTGA 2348
Db 656 GTTCTGTGTTATTTTCATTACACAACTTAATTTGTCTTGTAAATAGTAC-AGTGA 714

QY 2349 TCTTGGAGTGGGATTTCTTGGTAAATATCTTGGCACTTGAATGTCTCATGATTACATATG 2408
Db 715 TCTTGGAGTGGGATTTCTTGGTAAATATCTTGGCACTTGAATGTCTCATGATTACATATG 774

QY 2409 AAATCGCTTTGACATATCTTTAGACAGAAAAAGTAGCTGAGTGGGGGAAATATAGA 2468
Db 775 AAATCGCTTTGACATATCTTTAGACAGAAAAAGTAGCTGAGTGGGGGAAATATAGA 834

QY 2469 GCTTGTGACTTTAGGAGTAGCT 2493
Db 835 GC-TGTGTGACTTTAGGAGTAGGT 858

RESULT 7
CF272319 949 bp mRNA linear EST 13-AUG-2003
DEFINITION AGENCOURT_15193947 NIH_MGC_192 Homo sapiens cDNA clone
LOCUS IMAGE:30511964 5', mRNA sequence.
ACCESSION CF272319
VERSION CF272319.1 GI:33628231
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 949)
           NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Office of Cancer Genomics
COMMENT Contact: Daniela S. Gerhard, Ph.D.
           National Cancer Institute / NIH
           Bldg. 31 Rm10A07 Bethesda, MD 20892
           Email: cgapbs-remail.nih.gov
           Tissue Procurement: DCTD/DTP
           cDNA Library Preparation: Agencourt
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: NDAM571 row: h column: 21
           High quality sequence stop: 582.
FEATURES   Location/Qualifiers
           1..949
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="IMAGE:30511964"
           /tissue_type="Brain glioblastoma"
           /lab_host="DH10B (T1 phage-resistant)"
           /clone_lib="NIH_MGC_192"
           /note="Vector: pExpress1; Site 1: SmaI; Site 2: NotI; The
           library was constructed by reverse transcription of 1 ug
           mRNA using the oligo dr primer GCGGCGGCC(T)20 and an
           RNaseH + MMLV reverse transcriptase. Second strand
           synthesis was carried out by standard methods. The cDNA
           was size selected by agarose gel for > 1.2 kb, digested
           with Not I and directionally cloned into the vector
           Express-1 at the SmaI/NotI sites. DNA from the primary
           library was used for in vitro transcription from the T7
           promoter to produce biotinylated RNA transcripts. These
           biotinylated transcripts, along with blocking oligos to
           the poly-A, multiple cloning site and primer regions, were
           hybridized with single stranded circles produced by
           phageid production from the primary library to a Cot
           value of 10-20. Streptavidin/phenol extraction was utilized
           to remove DNA:RNA hybrids leaving un-hybridized single

```


QY 1560 CCGTCATCAAAAGTCCCTCTCATCATGTTCCAGTGAGAGCCAGCGATTGCTTTCTTC 1619
 |||||
 Db 360 CCGTCATCAAAAGTCCCTCTCATCATGTTCCAGTGAGAGCCAGCGATTGCTTTCTTC 419
 |||||
 QY 1620 CTGGCATAGTAACATTTCTTTGGAACATATGTTTCACCTTAATCATCTACCAATATCTGG 1679
 |||||
 Db 420 CTGGCATAGTAACATTTCTTTGGAACATATGTTTCACCTTAATCATCTACCAATATCTGG 479
 |||||
 QY 1680 AAGACCTGTCTTACTCAGACAGCACAGGTGTACAGAAGCAGACAGAAAGATCTTCCAGA 1739
 |||||
 Db 480 AAGACCTGTCTTACTCAGACAGCACAGGTGTACAGAAGCAGACAGAAAGATCTTCCAGA 539
 |||||
 QY 1740 TCAGCAGGAGAGCCCGGAGCCTCTGCTTCTCTACACTGGCATGCTGATGATGATGATGATG 1799
 |||||
 Db 540 TCAGCAGGAGAGCCCGGAGCCTCTGCTTCTCTACACTGGCATGCTGATGATGATGATGATG 599
 |||||
 QY 1800 CATGCCACATAGGCTTCTTCCACATCTGTTTGCACCTCGCTCATGATGGCTCGCTGCATC 1859
 |||||
 Db 600 CATGCCACATAGGCTTCTTCCACATCTGTTTGCACCTCGCTCATGATGGCTCGCTGCATC 659
 |||||
 QY 1860 TCCCTCAGTCCCAATTTAGTAGCAAGTGTCTCTGCGAGAGGCTGTCTATGTCTCTGG 1919
 |||||
 Db 660 TCCCTCAGTCCCAATTTAGTAGCAAGTGTCTCTGCGAGAGGCTGTCTATGTCTCTGG 719
 |||||
 QY 1920 CTGCCCAAGGACATCTCTGCGAGAGCATTTTGG 1954
 |||||
 Db 720 TTGCCAAGGACA---TCCTGCGAGAGCATTTGGG 751
 |||||

RESULT 9

BG497576

LOCUS

DEFINITION 601859469F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4073560 5', mRNA linear EST 27-MAR-2001
 mRNA sequence.

ACCESSION

BG497576

VERSION

BG497576.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

Fukuyama; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 767)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue procurement: DCTD/DTF

CDNA Library Prepared by: CLONETECH Laboratories, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL at:

<http://image.llnl.gov>

Plate: L1CM922 row: i column: 17

High quality sequence stop: 760.

Location/Qualifiers

1. .767

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4073560"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH_MGC 60"

/note="Organ: prostate; Vector: pDNR-LIB (Clontech);

Site1: SfII (ggccgctcgcc); Site2: SfII

(ggccattatggcc); Double-stranded cDNA was prepared from

cell line RNA. 5' and 3' adaptors were used in cloning as

follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and

3' adaptor sequence:

5'-ATTCTAGAGCGGAGCGCGGCATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.5

FEATURES

source

kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 26.7%; Score 685.2; DB 4; Length 767;
 Best Local Similarity 97.9%; Pred. No. 6.6e-151;
 Matches 735; Conservative 0; Mismatches 11; Indels 5; Gaps 4;

QY 1525 ACTTCAGGCATCGTACCCTCTTCTGATGAGAGCCCGTCATCACAAGAGTCCCTCTCAT 1584
 |||||
 Db 1 ACTTCAGGCATCGTACCCTCTTCTGATGAGAGCCCGTCATCACAAGAGTCCCTCTCAT 60
 |||||
 QY 1585 CATGTTCCAGTGAGAGGCGAGATGCTTCTTCTGCGCATAGTAAACATTTCTTTGGA 1644
 |||||
 Db 61 CATGTTCCAGTGAGAGGCGAGGANTGCTTCTTCTGCGCATAGTAAACATTTCTTTGGA 120
 |||||
 QY 1645 ACATATGTTTCACTTAATCACTACCAATATCTGAGAGACCTGTCTTACTCAGACAGCAC 1704
 |||||
 Db 121 ACATATGTTTCACTTAATCACTACCAATATCTGAGTACCTGTCTTACTCAGACAGCAC 180
 |||||
 QY 1705 CAGGTGTACAGAAGCAGACAGCAAGATCTTCAGATCAGCAGGAGACCCCGAGGCTCT 1764
 |||||
 Db 181 CAGGTGTACAGAAGCAGACAGCAAGATCTTCAGATCAGCAGGAGACCCCGAGGCTCT 240
 |||||
 QY 1765 GCTTCTCTACACTGGCATGCTGATGAGATCTGATGAGATGAGTCCACATTTGGCTTCTTCCA 1824
 |||||
 Db 241 GCTTCTCTACACTGGCATGCTGATGAGATCTGATGAGATGAGTCCACATTTGGCTTCTTCCA 300
 |||||
 QY 1825 TCTGTTGACCTGCTCATGATGGCTCGCTGATCTCCCTCAGTCCCAATTTCTAGTAGC 1884
 |||||
 Db 301 TCTGTTGACCTGCTCATGATGGCTCGCTGATCTCCCTCAGTCCCAATTTCTAGTAGC 359
 |||||
 QY 1885 CAAAGTGTCTCTGAGAGGCTGTCTATGTCTGCTGCTGCCAAGGACACTCTTCGAGAG 1944
 |||||
 Db 360 CAAAGTGTCTCTGAGAGGCTGTCTATGTCTGCTGCTGCCAAGGACACTCTTCGAGAG 418
 |||||
 QY 1945 CCATTTTGGTGAAGGAACACTTACAAAGAGGCAATGATTTGTGTCTGAGGCTCAGAG 2004
 |||||
 Db 419 CCATTTTGGTGAAGGAACACTTACAAAGAGGCAATGATTTGTGTCTGAGGCTCAGAG 478
 |||||
 QY 2005 CCCTTTTGTAGAGCTCTGATGCTCATTTATAAGACATTCAGCCCAAGTGTCCCAACTG 2064
 |||||
 Db 479 CCCTTTTGTAGAGCTCTGAGTCTGA-GTCAATATAAGACATTCAGCCCAAGTGTCCCAACTG 537
 |||||
 QY 2065 CAAATATACCAACCTTCTCTGAATATATTTTGGCTTATTTATATTTCTTTTCTTTTTC 2124
 |||||
 Db 538 CAAATATACCAACCTTCTCTGAATATATTTTGGCTTATTTATTTCTTTTCTTTTTC 597
 |||||
 QY 2125 TAAAGAATTGGCTCTGAATAGAAATGCACATTTTCCATCTGAATCTGATGCATATCATTTA 2184
 |||||
 Db 598 TAAAGTAT--GGTCTGAATAGAAATGCACATTTTCCATTTGGAATGGAATGCAITTCATTTA 655
 |||||
 QY 2185 GCCATCCAGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTCTCCTCAGCATAGGAGC 2244
 |||||
 Db 656 GCCAATCCAGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTCTCCTCAGCATAGGAGC 715
 |||||
 QY 2245 TATGATTCATTAATTAATAAGTGGAGTCAAAA 2275
 |||||
 Db 716 TATGATTCATTAATTAATAAGTGGAGTCAAAA 746
 |||||

RESULT 10

BG497758

LOCUS

DEFINITION

602543017F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4665484 5', mRNA sequence.

ACCESSION

BG497758

VERSION

BG497758.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

BG497758 759 bp mRNA linear EST 27-MAR-2001

602543017F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4665484 5',

mRNA sequence.

BG497758

BG497758.1 GI:13459275

EST.

Homo sapiens (human)

Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 759)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@email.nih.gov
 Tissue Procurement: DCTP/DTP
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLCM1471 row: a column: 05
 High quality sequence stop: 736.
 Location/Qualifiers
 1..759
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4665484"
 /issue_type="adenocarcinoma"
 /lab_host="NIH_MGC_60"
 /clone_lib="NIH_MGC_60"
 /notes="Organ: prostate; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccgctggcc); Site 2: SfiI (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGGCGCCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
 ORIGIN
 Query Match 26.2%; Score 674.8; DB 4; Length 759;
 Best Local Similarity 95.8%; Pred. No. 1.9e-148;
 Matches 724; Conservative 0; Mismatches 29; Indels 3; Gaps 3;
 QY 1207 GAAGACAGGGAAGAGCAGAACTTGAGGCCACAGTGTTCGTCGTCAGGAGTAAAT 1266
 DB 1 GAAGACAGGGAAGAGCAGAACTTGAGGCCACAGTGTTCGTCGTCAGGAGTAAAT 60
 QY 1267 GTCTACAGCGATGGAGATCTGCTGCCATGTGGTTGATGCCAGCCCTCCAGAGA 1326
 DB 61 GTCTACAGCGATGGAGATCTGCTGCCATGTGGTTGATGCCAGCCCTCCAGAGA 120
 QY 1327 AAAGGTCTGCAGACAGTATTAAAGCCCTAATCCAGATAGTATTAGTGAACCGTAGTTACT 1386
 DB 121 AAAGGTCTGCAGACAGTATTAAAGCCCTAATCCAGATAGTATTAGTGAACCGTAGTTACT 180
 QY 1387 CTGGCCAGGTGCCAGCTTACTAGATAGATAGTGTGTTGAACATCTACATCCACATT 1446
 DB 181 CTGGCCAGGTGCCAGCTTACTAGATAGATAGTGTGTTGAACATCTACATCCACATT 240
 QY 1447 TGTATACAGTGTCCAAATTTCTGTTCTACAGCATGTGTGTGCGAGAAACTGGAG 1506
 DB 241 TGTATGCAATGTTCCAAATTTCTGTTCTACAGCATGTGTGTGCGAGAAACTGGAG 300
 QY 1507 ACCAGGCATCTTAATTTTACTTACGCATCGTACCCCTCTTCTGACTGATGAGACCCGTCAT 1566
 DB 301 ACCAGGCATCTTAATTTTACTTACGCATCGTACCCCTCTTCTGACTGATGAGACCCGTCAT 360
 QY 1567 CACAAGGTCCTCTCATGTTCCAGTGAGAGCCAGGATGCTTTCTTCTCTGCGAT 1626
 DB 361 CACAAGGTCCTCTCATGTTCCAGTGAGAGCCAGGATGCTTTCTTCTCTGCGAT 420
 QY 1627 AGTAAACATTTTCTTGGACATATGTTTCACTTAATCACTACCAAAATATCTGGAAGACCT 1686

DB 421 AGTAAACATTTTCTTGGACATATGTTTCACTTAATCACTACCAAAATATCTGGAAGACCT 480
 QY 1687 GTCTTACTCAGACAGACACAGGTGTACAGAACAGCAGACAGATCTTCCAGATCAGCAG 1746
 DB 481 GTCTTACTCAGACAGACACAGGTGTACAGAACAGCAGACAGATCTTCCAGATCAGCAG 540
 QY 1747 GGAGACCCCGGAGCCTCTGCTTCTCTACACT-GGCATGCTGATGAGATCGTGACATGCC 1805
 DB 541 GGAGACCCCGGAGCCTCTGCTTCTCTACACTGCTGGGATGCTGATGAGATCGTGACATGCC 600
 QY 1806 CACATTGGCTTCTTCCACATCTGGTTGCACTCG-TCATGATGGGCTCGTGTGATCTCCCT 1864
 DB 601 CACATTGGCTTCTTCCACATCTGGTTGCACTCGTTCATGATGGGCTCGTGTGATCTCCCT 660
 QY 1865 CAGTCCCAAAATCTAGTACGCAAGTGTCTGAGAGGCTGTCTAGTGTCTCTGCGC-TGC 1923
 DB 661 CAAGTCCCAAAATCTAGTACGCAAGTGTCTGAGAGGCTGTCTAGTGTCTCTGCGCTTGC 720
 QY 1924 CCAAGGGACACTCTCGCAGAGCCATTTTGGGTAAG 1959
 DB 721 CCAAGGACAACTCTCGCAGACCACTTAGGTAAG 756
 RESULT 11
 BUI183182
 LOCUS
 DEFINITION
 AGENCOURT_7952368 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:6108122
 5', mRNA sequence.
 ACCESSION
 BUI183182
 VERSION
 BUI183182.1 GI:22697166
 KEYWORDS
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 759)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@email.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLCM2351 row: o column: 03
 High quality sequence stop: 572.
 Location/Qualifiers
 1..759
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6108122"
 /issue_type="from acute myelogenous leukemia"
 /lab_host="DH10B (TL phage-resistant)"
 /clone_lib="NIH_MGC_55"
 /notes="Organ: bone marrow; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccgctggcc); Site 2: SfiI (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGGCGCCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
 ORIGIN


```

Db      539 TGATCTTGCTGCTGAGGCTCAGAGCCCTTTGATAGGCTTCTGA-GTCATATATAAGAC 597
QY      2041 ATTCAAGCCAGATGCTCAACTGCAGAAATATACCAACCTTCTCTGAATATATTTGCTT 2100
Db      598 ATTCAAGCCAGATGCTCAACTGCAGAAATATATACCAACTTCTCTGAATATATTTGCTT 657
QY      2101 ATTATATTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2149
Db      658 ATTATATTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 705

RESULT 13
BI551715
LOCUS      603197411F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5276907 5',
DEFINITION mRNA sequence.
ACCESSION BI551715
VERSION    BI551715.1 GI:15439040
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 833)
            NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LHAM11699 row: e column: 04
            High quality sequence stop: 726.
FEATURES
            source
            1..833
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5276907"
            /tissue_type="hippocampus"
            /lab_host="DH10B"
            /clone_lib="NIH_MGC_95"
            /notes="Organ: brain; Vector: pBluescriptR (modified
            pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
            (gtcgag); Oligo-dt primed using primer
            5'-TTTTTTTTTTTTTTTNN-3', size-selected for average
            insert size 2.5 kb and normalized to ROT 5. This is a
            primary library enriched for full-length clones and
            constructed using the Cap-trapper method (Carninci, in
            preparation). Library constructed by M. Brownstein
            (NIH/NHGRI, National Institutes of Health). Note: this
            is a NIH_MGC Library."
ORIGIN
Query Match 25.1%; Score 646.6; DB 4; Length 833;
Best Local Similarity 95.4%; Pred. No. 9.3e-142;
Matches 795; Conservative 0; Mismatches 24; Indels 14; Gaps 12;
QY      272 CCGGACGCGCGCTCGGGGCGCGGCTCGGGCGCGCGGCTGCGACCGCGCTCGTGAC 331
Db      3 CCGGACGCGCGCTCGGGGCGCGGCTCGGGCGCGCGGCTGCGACCGCGCTCGTGAC 61
QY      332 CAGCTGCGCGCGCGCGCTCGAGCGGTGCCAGTGTCTCAGGCTGCTGTCTACTGCGCG 391
Db      62 CAGCTGCGCGCGCGCGCTCGAGCGGTGCCAGTGTCTCAGGCTGCTGTCTACTGCGCG 121
QY      392 GCGCGCGACGCGCGCGCGCGACAGCAAGGCTTCTGCTGCGCGACCCCTCGATGACCT 451

```

```

Db      122 GCGCGCCAGGCGCGCGCGCAGCAAGGCTTCTTGCTGCGCGACCCCTCGATGACCT 181
QY      452 GACACCCGCGAAGCGCTGCTCGAGCTGCTGGCGGCTGCGAGGAGCAGCAGCCCGCGAC 511
Db      182 GACACCCGCGAAGCGCTGCTCGAGCTGCTGGCGGCTGCGAGGAGCAGCAGCCCGCGAC 241
QY      512 TTGGGCGAGTTCGAGCGCGACCCGCGCGCGCAGCTGTGGCAGCGCTCTGGGAGGTGCAA 571
Db      242 TTGGGCGAGTTCGAGCGCGACCCGCGCGCGCAGCTGTGGCAGCGCTCTGGGAGGTGCAA 301
QY      572 GACGCGAGCGGCTCGAGGTGGGCTGCGCACAGGTGCTGCCCGT-CCGCGAGCCCGCGCT 630
Db      302 GACGCGAGCGGCTCGAGGTGGGCTGCGCACAGGTGCTGCCCGTCCCGGAGCCCGCGCT 361
QY      631 GACCCGCGGTGGCGCAGACTTCCCGAGTTCCTGGTCTTCCCGGACCGGAGAGCCCGCG 690
Db      362 GACCCGCGGTGGCGCAGACTTCCCGAGTTCCTGGTCTTCCCGGACCGGAGAGCCCGCG 421
QY      691 GCGCGTTTTGGAGGAGTGTACCTTCTTTATTCCTGAAGCCCGGCGAGTGTTCGACTGGT 750
Db      422 GCGCGTTTTGGAGGAGTGTACCTTCTTTATTTCTGAAGCCCGGCGAGTGTTCGACTGGT 481
QY      751 CGACCAAGTCCCAAAACAGATCCAGAAAGTTCAGGTTGTTGCCATCGAAGGACT 810
Db      482 CGACCAAGTCCCAAAACAGATCCAGAAAGTTCAGGTTGTTGCCATCGAAGGACT 540
QY      811 GGATGCCAGGGT-AAAACCAAGGTGACCC-AGTCAGTGCAGATTCACCTTAAGGCTGTC 868
Db      541 GGATGCCAGGGTAAAACCAAGGTGACCC-AGTCAGTGCAGATTCACCTTAAGGCTGTC 600
QY      869 CTCTTAAAGTCACCAACCTCTTGCA-TTGCGCAGTGGAGGAAGATCTTTGATGATGAACC 927
Db      601 CTCTTAAAGTCACCAACCTCTTGCA-TTGCGCAGTGGAGGAAGATCTTTGATGATGAACC 660
QY      928 AACTATCATPAGAAGAGCTTTTACTCTTTTGGGCAATATATTTGGGCT-CCGAATAG 986
Db      661 AGTATCATTAGAAGAGCTTTTACTCTTTGGGCCACTTATATTTGGGCTCCCGAAATAG 720
QY      987 CTAAGAACTGCGCAA-TCTCCTGTGATTGTA-GACAGTACTGCGCAGCAGC-GGCC 1042
Db      721 CTAAGAACTGCGCAA-TCTCCTGTGATTGTA-GACAGTACTGCGCAGCAGC-GGCC 780
QY      1043 ACCTATGCCA-TAGCCACTGAGGTGA--GTGGGGGTCTCCAGCACTTGGCCCC 1092
Db      781 ACCTATGCCATTAGGCACCTGAGTGAAGTGGGGGTATCCAGAACCGGACCC 833

RESULT 14
BG778145 844 bp mRNA linear EST 15-MAY-2001
LOCUS     60266482F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4806117 5',
DEFINITION mRNA sequence.
ACCESSION BG778145
VERSION    BG778145.1 GI:14048462
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 844)
            NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTF
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov

```

Plate: L16M1655	row: d	column: 22
High quality sequence stop: 631.		
Location/Qualifiers		
1. .844		
/organism="Homo sapiens"		
/mol_type="mRNA"		
/db_xref="taxon:9606"		
/clone="IMAGE:4806117"		
/tissue_type="adenocarcinoma"		
/lab_host="DH10B (T1 phage-resistant)"		
/clone_lib="NIH_MGC_60"		
/note="Organ: prostate; Vector: pDNR-LIB (Clontech); Site1: SfII (ggcgctcgccg); Site2: SfII (ggccattatggc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGGCCATTATGGCC-3' and 3' adaptor sequence:		
5'-ATTGTAGAGCGGCGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.5 Kb (range 0.9-4.0 Kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."		
ORIGIN		
Query Match	24.2%;	Score 621.8; DB 4; Length 844;
Best Local Similarity	98.9%;	Pred.No. 6.9e-136;
Matches 626;	Conservative 0;	Mismatches 7; Indels 0; Gaps 0;
QY 1018	AGACAGTACTGSCACAGCAGCGCCACTATGCCATAGCCACCTCAGTGAGTGGGGTCT	1077
Db 1	AGACAGTACTGGCAGCAGCGGCCACTATGCCATAGCCACCTCAGTGAGTGGGGTCT	60
QY 1078	CCAGCACTGCCCGCCAGCCCATCACCTGTGTACAGTGGCCAGAGACCTGCTCAAAAC	1137
Db 61	CCAGCACTGCCCGCCAGCCCATCACCTGTGTACAGTGGCCAGAGACCTGCTCAAAAC	120
QY 1138	TGACCTTATCCTGCTGCTACTGTGAGTCTTGAGGAGAGGTTCGAGAGGCTCGAGGGCCG	1197
Db 121	TGACCTTATCCTGCTGCTACTGTGAGTCTTGAGGAGAGGTTCGAGAGGCTCGAGGGCCG	180
QY 1198	GGGCATGGGAAGACAGGGAAGACAGAACTTGAGGCCAACAGTGTGTTTGGTCAAAA	1257
Db 181	GGGCATGGGAGAGACAGGAGAGAGCAGAACTTGAGGCCAACAGTGTGTTTGGTCAAAA	240
QY 1258	GGTAGAAATGCTCTACACAGCGATGGAGAAATCCTGCTGCCATGTGTTGATGCCAGCCC	1317
Db 241	GGTAGAAATGCTCTACACAGCGATGGAGAAATCCTGCTGCCATGTGTTGATGCCAGCCC	300
QY 1318	CTCCAGAGAAAGGTCTGCAGACAGTATTAAAGCCTAATCCAGAAATGTTTTAGTGAACC	1377
Db 301	CTCCAGAGAAAGGTCTCTGCAGACGGTATTAAAGCCTAATCCAGAAATGTTTTAGTGAACC	360
QY 1378	GTAGTTACTCTGGCCAGGTGCCACGTCTAACTAGATTAGATTGTTTGAACAATCTACA	1437
Db 361	GTAGTTACTCTGGCCAGGTGCCACGTCTAACTAGATTAGATTGTTTGAACAATCTACA	420
QY 1438	TCCACATTTGTTATGCAAGTGTCCCAAAATTCGTGTTCAAGACATGTTGTGTGGCAGA	1497
Db 421	TCCACATTTGTTATGCAAGTGTCCCAAAATTCGTGTTCAAGACATGTTGTGTGGCAGA	480
QY 1498	AAACTGGACACAGGCATCTTAATTTTACTTCAGGCATCGTACCTCTTCGACTGATGG	1557
Db 481	AAACTGGAGACAGGCATCTTAATTTTACTTCAGGCATCGTACCTCTTCGACTGATGG	540
QY 1558	ACCCGTCATCAAAAAGTCCCTCTCATCATGTTTCCAGTGAGAGCCAGCGATTGCTTTCT	1617
Db 541	ACCCGTCATCAAAAAGTCCCTCTCATCATGTTTCCAGTGAGAGCCAGCGATTGCTTTCT	600
QY 1618	TCCTGGCATAGTAAACATTTTCTTGGAAACAT	1650
Db 601	TCCTGGCATAGTAAACATTTTCTTGGAAACAT	633

```

|||||
Db      242  TGACATGCCACATTGGCTTCTTCCACATCTGGTTGCACTCGTCAATGATGGGTCGCTGC 301
QY      1857  ATCTCCCTCAGTCCCAAAATCTAGTAGCCAAAGTGTCTCTGCAGAGGCTGTCTATGTGTCC 1916
Db      302  ATCTCCCTCAGTCCCAAAATCTAGTAGCCAAAGTGTCTCTGCAGAGGCTGTCTATGTGTCC 360
QY      1917  TGGCTGCCCAAGGGACACTCTCTGCAGAGGCAATTTTGGGTAAGGAACACTTACAAAGAAG 1976
Db      361  TGGCTGCCCAAGGGACACTCTCTGCAGAGGCAATTTTGGGTAAGGAACACTTACAAAGAAG 419
QY      1977  GCATTGATCTTGTGCTGTGGGCTCAGAGCCCTTTTGATAGGCTTCTGATGCTCATTCTATAA 2036
Db      420  GCATTGATCTTGTGCTGTGGGCTCAGAGCCCTTTTGATAGGCTTCTGATGCTCATTCTATAA 479
QY      2037  AGACATTTCAAGCCCAAGATGCTCCCACTGCAAAATATACCAACCTTCTCTGAATTTATATTT 2096
Db      480  AGACATTTCAAGCCCAAGATGCTCCCACTGCAAAATATACCAACCTTCTCTGAATTTATA-TTT 538
QY      2097  GCTTATTTATATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2156
Db      539  GCTTATTTATATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 598
QY      2157  TCCATCTGAACTGGATGCTATCATTTAGCCATCCAGTAATTTATTTATTTATTTATTTATTTAT 2216
Db      599  TTCCATTTGAACTGGATGCTATCATTTAGCCATCCAGTAATTTATTTATTTATTTATTTATTTAT 657
QY      2217  ACAT-AATATGTTTCTCTCAGCATA-GGAGCTATGATTCTATTAATTTAAAGTGGAGTCAAA 2274
Db      658  ACATAAATAATGTTTCTCTCAGCATAAGGAGCTATGATTCTTAAACGTTGGAGTCAAA 717
QY      2275  A 2275
Db      718  A 718
```

Search completed: October 22, 2004, 20:20:39
Job time : 7932 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 22, 2004, 18:08:37 ; Search time 1219 Seconds

(without alignments)
10806.892 Million cell updates/sec

Title: US-10-681-223-1

Perfect score: 2571

Sequence: 1 cggggcggaggcggcgtcg.....aggaggggcttttgcaccc 2571

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3407233 seqs, 2561960514 residues

Total number of hits satisfying chosen parameters: 6814466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2571	100.0	2571	14	US-10-277-032-1
2	2571	100.0	2571	16	US-10-681-223-1
3	2276	88.5	2856	15	US-10-305-810-6
4	2125.2	82.7	2645	13	US-10-044-090-720
5	1130.4	44.0	20966	14	US-10-277-032-3
6	1130.4	44.0	20966	16	US-10-681-223-3
7	1035.4	40.3	1579	16	US-10-264-237-104
8	303.8	11.8	507	9	US-09-867-701-2642
9	219.4	8.5	497	9	US-09-783-590-5564
10	149.8	5.8	603	16	US-10-240-425-827
11	126.2	4.9	390	9	US-09-783-590-11496
12	79.8	3.1	2118	15	US-10-156-761-7362
					Sequence 1, Appli
					Sequence 1, Appli
					Sequence 6, Appli
					Sequence 720, App
					Sequence 3, Appli
					Sequence 3, Appli
					Sequence 104, App
					Sequence 2642, Ap
					Sequence 5564, Ap
					Sequence 827, App
					Sequence 11496, A
					Sequence 7362, Ap

13	79.8	3.1	9025608	15	US-10-156-761-1	Sequence 1, Appli
14	76	3.0	1413	10	US-09-953-348-51	Sequence 51, Appl
15	76	3.0	1413	10	US-10-267-255-51	Sequence 51, Appl
16	76	3.0	53500	15	US-09-953-348-76	Sequence 76, Appl
17	76	3.0	53500	15	US-10-267-255-76	Sequence 76, Appl
18	75.8	2.9	667	17	US-10-767-701-4633	Sequence 4633, Ap
19	75	2.9	1092	17	US-10-767-701-9733	Sequence 9733, Ap
c	74	2.9	1860	15	US-10-156-761-1558	Sequence 1558, Ap
c	74	2.9	9025608	15	US-10-156-761-1	Sequence 1, Appli
22	73	2.8	1437	15	US-10-156-761-5988	Sequence 5988, Ap
23	71.8	2.8	1041	17	US-10-437-963-8597	Sequence 8597, Ap
c	71.6	2.8	2561	9	US-09-976-740-48	Sequence 48, Appl
c	71.6	2.8	2561	13	US-10-023-523-48	Sequence 48, Appl
c	71.6	2.8	2561	13	US-10-023-523-48	Sequence 48, Appl
c	71.6	2.8	2561	16	US-10-616-187-48	Sequence 48, Appl
c	71.6	2.8	2561	16	US-10-671-242-48	Sequence 48, Appl
c	71.6	2.8	2561	16	US-10-671-242-48	Sequence 48, Appl
c	71	2.8	1388	17	US-10-767-701-14298	Sequence 14298, A
c	71	2.8	1734	15	US-10-156-761-6476	Sequence 6476, Ap
c	70	2.7	15738	15	US-10-329-079-46	Sequence 46, Appl
c	70	2.7	61944	15	US-10-329-079-34	Sequence 34, Appl
32	68.8	2.7	909	11	US-09-758-759-156	Sequence 156, App
34	68.8	2.7	109519	11	US-09-758-759-1	Sequence 1, Appli
c	68.6	2.7	3459	15	US-10-369-493-31552	Sequence 31552, A
c	67.2	2.6	1618	15	US-10-104-047-1463	Sequence 1463, Ap
37	66.8	2.6	1946	17	US-10-437-963-49955	Sequence 49955, A
38	66.4	2.6	1843	17	US-10-437-963-22190	Sequence 22190, A
c	66.2	2.6	2520	10	US-09-873-367C-159	Sequence 159, App
c	66	2.6	924	15	US-10-156-761-1189	Sequence 1189, Ap
c	66	2.6	1521	15	US-10-156-761-3755	Sequence 3755, Ap
c	66	2.6	1725	17	US-10-324-967-35	Sequence 35, Appl
c	65.6	2.6	763	17	US-10-437-963-49370	Sequence 49370, A
c	65.6	2.6	3957	15	US-10-200-562-193	Sequence 193, App
45	65.6	2.6	3957	15	US-10-237-551-193	Sequence 193, App

ALIGNMENTS

RESULT 1
US-10-277-032-1
; Sequence 1, Application US/10277032
; Publication No. US20030087294A1
; GENERAL INFORMATION:
; APPLICANT: Ming-Hui WEI
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001305 DIV
; CURRENT APPLICATION NUMBER: US/10/277,032
; CURRENT FILING DATE: 2002-10-22
; PRIOR FILING DATE: 09/984,880
; PRIOR APPLICATION NUMBER: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2571
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-277-032-1

Query Match	100.0%	Score	2571	DB	14	Length	2571
Best Local Similarity	100.0%	Pred. No.	0				
Matches	2571	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
QY	1	CGGGGCGGAGGCGGCGTGGTGGGCGCCATGGCTTGGCCGCGCGGCTCTCTGGCGCG	60				
Db	1	CGGGGCGGAGGCGGCGTGGTGGGCGCCATGGCTTGGCCGCGCGGCTCTCTGGCGCG	60				
QY	61	GCACTGTCTGGGCGCGCTCTCTGGGCGGCGCGGGGTCTCGCTGGGGCCATGGCTCCGCC	120				
Db	61	GCACTGTCTGGGCGCGCTCTCTGGGCGGCGCGGGGTCTCGCTGGGGCCATGGCTCCGCC	120				
QY	121	GTGCGCTTCGCTCTGGAGCTTCCGACCTGACCCCTGGCTCACTTCGCCCTAGGCGCCGA	180				


```
QY 2341 CAAGTGGATCTTGGAGTGGGATTTCTTGGTAAATTAATCTTGACATTTGAATGTCATGAT 2400
Db 2341 CAAGTGGATCTTGGAGTGGGATTTCTTGGTAAATTAATCTTGACATTTGAATGTCATGAT 2400
QY 2401 TACATATGAATCGCTTTGACATATCTTTAGACAGAAAAAAGTAGCTGAGTGAGGGGAA 2460
Db 2401 TACATATGAATCGCTTTGACATATCTTTAGACAGAAAAAAGTAGCTGAGTGAGGGGAA 2460
QY 2461 ATTATAGAGCTTGTGTGATCTTTAGGAGTAGCTGCTTTATACACATACTCAAGCCCTG 2520
Db 2461 ATTATAGAGCTTGTGTGATCTTTAGGAGTAGCTGCTTTATACACATACTCAAGCCCTG 2520
QY 2521 AAGCCTTGATGCTCCTGACGCTGCGACATAAGGAGGGGCTTTTGACCC 2571
Db 2521 AAGCCTTGATGCTCCTGACGCTGCGACATAAGGAGGGGCTTTTGACCC 2571

RESULT 2
US-10-681-223-1
; Sequence 1, Application US/10681223
; Publication No. US20040081999A1
; GENERAL INFORMATION:
; APPLICANT: Ming-Hui WEI
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001305 DIV-11
; CURRENT APPLICATION NUMBER: US/10/681,223
; PRIOR FILING DATE: 2003-10-09
; PRIOR APPLICATION NUMBER: 10/277,032
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/984,880
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2571
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-681-223-1

Query Match 100.0%; Score 2571; DB 16; Length 2571;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGGCGGAGGGCGGCTCGCTGAGCGGCCCATGGCTTTCGCCCGCGGCTCTCGCGCG 60
Db 1 CGGGGCGGAGGGCGGCTCGCTGAGCGGCCCATGGCTTTCGCCCGCGGCTCTCGCGCG 60
QY 61 GGCACCTGTCGGGGCGGCTGCTCGGGGGCGGGGGTCTGGCTGGGGCCCATGGCTCCGCC 120
Db 61 GGCACCTGTCGGGGCGGCTGCTCGGGGGCGGGGGTCTGGCTGGGGCCCATGGCTCCGCC 120
QY 121 GTGCGGCTTCGCTCGAGCTTCCGACCTGACCTCGGCTGCTTTCGCCCGCGGCTCTCG 180
Db 121 GTGCGGCTTCGCTCGAGCTTCCGACCTGACCTCGGCTGCTTTCGCCCGCGGCTCTCG 180
QY 181 CGCCCCCGCGCAGCAGACGCCCGGACCCCGGCTGCGGGGCTGTTGGGGCCCCCGGGA 240
Db 181 CGCCCCCGCGCAGCAGACGCCCGGACCCCGGCTGCGGGGCTGTTGGGGCCCCCGGGA 240
QY 241 GGCAGCTACTGCTGTGCTGCGGCGGCGGGGCTGCGGCGGCGGCGGCGGCGGCTCCG 300
Db 241 GGCAGCTACTGCTGTGCTGCGGCGGCGGGGCTGCGGCGGCGGCGGCGGCGGCTCCG 300
QY 301 GCGGCGGCGGCTGACACGAGGCTGCTGACACGAGCTGCGCGGCGGCGGCTTCCAGCGGTG 360
Db 301 GCGGCGGCGGCTGACACGAGGCTGCTGACACGAGCTGCGCGGCGGCGGCTTCCAGCGGTG 360
QY 361 CCAGCTGCTCAGGCTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db 361 CCAGCTGCTCAGGCTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
```

```
QY 421 CTTCTGCTGCGGACCCCTCGATGACCTGACACACCCCGGCAAGCGCTGCTCGAGCTGCT 480
Db 421 CTTCTGCTGCGGACCCCTCGATGACCTGACACACCCCGGCAAGCGCTGCTCGAGCTGCT 480
QY 481 GGGCGCTGCCAGGAGGACACCGCGCACCTTTGGCGAGTTTCGAGCGCGACCCCGCGCG 540
Db 481 GGGCGCTGCCAGGAGGACACCGCGCACCTTTGGCGAGTTTCGAGCGCGACCCCGCGCG 540
QY 541 CCAGCTGTGGCAGCGCTCTGGGAGGTCAAGACGCGCAGCGCGCTGCAAGTGGGTGCGC 600
Db 541 CCAGCTGTGGCAGCGCTCTGGGAGGTCAAGACGCGCAGCGCGCTGCAAGTGGGTGCGC 600
QY 601 ACAGTGTGTCGCGCTCCCGGAGCCCGCTGCAACCGCGGTGGTGGTCCAGACTTGGCCAGTTC 660
Db 601 ACAGTGTGTCGCGCTCCCGGAGCCCGCTGCAACCGCGGTGGTGGTCCAGACTTGGCCAGTTC 660
QY 661 CGTGTGCTTCCCGGACCGGGAGCGCGCGGCTTTTGGAGGAGTGTACCTCTCTTTAT 720
Db 661 CGTGTGCTTCCCGGACCGGGAGCGCGCGGCTTTTGGAGGAGTGTACCTCTCTTTAT 720
QY 721 TCCTGAAGCCCGGCGAGTGTGACCTGCTGACCTGCTGACCTGCTGCTGCTGCTGCTGCT 780
Db 721 TCCTGAAGCCCGGCGAGTGTGACCTGCTGACCTGCTGACCTGCTGCTGCTGCTGCTGCT 780
QY 781 AAAGTTCAGGTTGTTGCCATCGAAGGACTGATGCGCAGCGGTAAACACCGGTGACCCA 840
Db 781 AAAGTTCAGGTTGTTGCCATCGAAGGACTGATGCGCAGCGGTAAACACCGGTGACCCA 840
QY 841 GTCAGTGGCAGATTACATTAAGGCTGCTCTTTAAAGTCAACACCTCTTTGATTTGGCCA 900
Db 841 GTCAGTGGCAGATTACATTAAGGCTGCTCTTTAAAGTCAACACCTCTTTGATTTGGCCA 900
QY 901 GTGGAGGAGATCTTTGATGATGAACCAACTATCATTAAGAGAGCTTTTACTCTTTGGG 960
Db 901 GTGGAGGAGATCTTTGATGATGAACCAACTATCATTAAGAGAGCTTTTACTCTTTGGG 960
QY 961 CAATTATATTGCGCTCCGAAATAGCTAAAGAACTGCAAAATCTCTCTGTGATTTAGTA 1020
Db 961 CAATTATATTGCGCTCCGAAATAGCTAAAGAACTGCAAAATCTCTCTGTGATTTAGTA 1020
QY 1021 CAGGTACTGGCAGACGCGCCACCTATGCCATAGCCTAGGAGTGGGGTCTCCA 1080
Db 1021 CAGGTACTGGCAGACGCGCCACCTATGCCATAGCCTAGGAGTGGGGTCTCCA 1080
QY 1081 GCACCTGCCCCAGCCCATCACCTGTGTACAGTGGCCAGAGGACCTGCTCAAACTGA 1140
Db 1081 GCACCTGCCCCAGCCCATCACCTGTGTACAGTGGCCAGAGGACCTGCTCAAACTGA 1140
QY 1141 CTTATCTCTGCTGCTCACTGTGAGTCTCTGAGAGAGTTGCAAGGCTGCGGGCGGGG 1200
Db 1141 CTTATCTCTGCTGCTCACTGTGAGTCTCTGAGAGAGTTGCAAGGCTGCGGGCGGGG 1200
QY 1201 CATGAGAAAGACACGAGGAAGAGCACTTCAGAGCCCAACAGTGTGTTTCGTCAAAAGT 1260
Db 1201 CATGAGAAAGACACGAGGAAGAGCACTTCAGAGCCCAACAGTGTGTTTCGTCAAAAGT 1260
QY 1261 AGAAATGCTCTACAGCGGATGGAGAACTCTGGCTGCCATGTGTTGATGCGAGCCCTTC 1320
Db 1261 AGAAATGCTCTACAGCGGATGGAGAACTCTGGCTGCCATGTGTTGATGCGAGCCCTTC 1320
QY 1321 CAGAGAAAGGTCCTGCGACAGTAAAGCTTAATCCAGAAATAGTTTTAGTGAACCGTA 1380
Db 1321 CAGAGAAAGGTCCTGCGACAGTAAAGCTTAATCCAGAAATAGTTTTAGTGAACCGTA 1380
QY 1381 GTTACTCTGGCAGGTGCCAGTCTAACTAGATTAGTGTGTTTGAACATCTACATCC 1440
Db 1381 GTTACTCTGGCAGGTGCCAGTCTAACTAGATTAGTGTGTTTGAACATCTACATCC 1440
QY 1441 ACCATTGTTATGACAGTGTCCCAAAATTTCTGTTCTAAGCATGTGTGTGGCAGAAAA 1500
Db 1441 ACCATTGTTATGACAGTGTCCCAAAATTTCTGTTCTAAGCATGTGTGTGGCAGAAAA 1500
```


Db 301 TACTGCCGGCGCCAGGCGCGGCGCACACAGCAAGGCTTCTCTGCTGCCGACCCCTG 360
QY 443 GATGACCTGACACCGCGCAAGCGCTGCTCGAGCTGCTGGCGCCTGCGCAGAGGACCA 502
Db 361 GATGACCTGACACCGCGCAAGCGCTGCTCGAGCTGCTGGCGCCTGTCAGAGGACCA 420
QY 503 CGCCCGCACTTGGCGAGTTCGAGGCGCGACCGCGCGCGCAGCTGCGCAGCGCCTG 562
Db 421 CGCCCGCACTTGGCGAGTTCGAGGCGCGACCGCGCGCGCAGCTGCGCAGCGCCTG 480
QY 563 GAGGTCAAGACGCGCAGCGCGCTGCGAGGTGGGTGGGCAAGGTCCAGGTTCGCGGAG 622
Db 481 GAGGTCAAGACGCGCAGCGCGCTGCGAGGTGGGTGGGCAAGGTTCGCGGAG 540
QY 623 CCCCCTGACACCGGTGGTGCAGACTTGCAGTTCGCCAGTTCCTGCGCGAGCGGGA 692
Db 541 CCCCCTGACACCGGTGGTGCAGACTTGCAGTTCGCCAGTTCCTGCGCGAGCGGGA 600
QY 683 GCGCGCGCGGCGCTTTTGGAGGAGTGTACCTCTCTTTATTTCTGAAGCGCGGCGAGTGT 742
Db 601 GCGCGCGCGGCGCTTTTGGAGGAGTGTACCTCTCTTTATTTCTGAAGCGCGGCGAGTGT 660
QY 743 GACCTGGTGCAGCAGTGCACCAACAGATCCAGAAAGAAAGTTCAGGTTCGCGATC 802
Db 661 GACCTGGTGCAGCAGTGCACCAACAGATCCAGAAAGAAAGTTCAGGTTCGCGATC 720
QY 803 GAAGGAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 859
Db 721 GAAGGAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 780
QY 860 AAGGCTGTCTCTTAAAGTCAACCACTCTTGCATTTGGCAGTGGAGGAAGATCTTTGAT 919
Db 781 AAGGCTGTCTCTTAAAGTCAACCACTCTTGCATTTGGCAGTGGAGGAAGATCTTTGAT 840
QY 920 GATGAACCAACTATCATTAAGAGAGCTTTTACTCTTTGGGCAATATATTTGTGGCTCC 979
Db 841 GATGAACCAACTATCATTAAGAGAGCTTTTACTCTTTGGGCAATATATTTGTGGCTCC 900
QY 980 GAAATAGCTTAAGAAATCTGCCAATCTCTGTGATTTGTAGACAGGTACTGGCAGCAG 1039
Db 901 GAAATAGCTTAAGAAATCTGCCAATCTCTGTGATTTGTAGAC- - - - -GGCAGCAGCAG 954
QY 1040 GCCACCTATGCCATAGCAGTGAAGTGGGGTCTCCAGCAGCCTGCCCGAGCCCAT 1099
Db 955 GCCACCTATGCCATAGCAGTGAAGTGGGGTCTCCAGCAGCCTGCCCGAGCCCAT 1014
QY 1100 CACCTGTGTACAGTGGCCAGAGGACCTGCTCAAACTGACCTTATCCTGCTGCTCACT 1159
Db 1015 CACCTGTGTACAGTGGCCAGAGGACCTGCTCAAACTGACCTTATCCTGCTGCTCACT 1074
QY 1160 GTGAGTCTGAGGAGGTTGCAGAGCTGCAGGCGCGGCGGCGATGAGAGACCGGAA 1219
Db 1075 GTGAGTCTGAGGAGGTTGCAGAGCTGCAGGCGCGGCGGCGATGAGAGACCGGAA 1134
QY 1220 GAAGCAGAACTTGAGGCCAACTGAGTGTGTTGTCFAAAGTGAAGATGCTTACAGCGG 1279
Db 1135 GAAGCAGAACTTGAGGCCAACTGAGTGTGTTGTCFAAAGTGAAGATGCTTACAGCGG 1194
QY 1280 ATGGAGAATCTGGCTGCCATGTTGATGCCAGCCTCCAGAGAAAGGTCTGCGAG 1339
Db 1195 ATGGAGAATCTGGCTGCCATGTTGATGCCAGCCTCCAGAGAAAGGTCTGCGAG 1254
QY 1340 ACAGTATTAGCTTAATCCAGATAGTTTTAGTGAACCGTAGTACTCTGGCCAGGTGCC 1399
Db 1255 ACGGTATTAGCTTAATCCAGATAGTTTTAGTGAACCGTAGTACTCTGGCCAGGTGCC 1314
QY 1400 ACGTCTAACTAGATTAGATTGTTTGAACATCTACATCCACATTTCTTATGCACTGT 1459
Db 1315 ACGTCTAACTAGATTAGATTGTTTGAACATCTACATCCACATTTCTTATGCACTGT 1374
QY 1460 TCCCAATTTCTGTTCTACAAGCATGTTGTGGCAGAAATCTGGAGACCGCATCTTA 1519
Db 1375 TCCCAATTTCTGTTCTACAAGCATGTTGTGGCAGAAATCTGGAGACCGCATCTTA 1434

QY 1520 ATTTTACTTCAGCATCGTACCCCTCTTCTGACGTGATGGACCCCTCATCAAAAGTCCCT 1579
Db 1435 ATTTTACTTCAGCATCGTACCCCTCTTCTGACGTGATGGACCCCTCATCAAAAGTCCCT 1494
QY 1580 CTCATCATGTTCCAGTGAGAGGCGCGGATGCTTCTTCTCGGCATAGTAAACATTTTC 1639
Db 1495 CTCATCATGTTCCAGTGAGAGGCGCGGATGCTTCTTCTCGGCATAGTAAACATTTTC 1554
QY 1640 TTGGAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGTCTTACTCAGAC 1699
Db 1555 TTGGAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGTCTTACTCAGAC 1614
QY 1700 AGCACCAGGTGTACAGAGCAGACAGATCTTCCAGATCAGAGGAGAGACCCGGAG 1759
Db 1615 AGCACCAGGTGTACAGAGCAGACAGATCTTCCAGATCAGAGGAGAGACCCGGAG 1674
QY 1760 CTTCTGCTTCTCTACACTGGCATGCTGATGATGATGATGATGATGATGATGATGATGAT 1819
Db 1675 CTTCTGCTTCTCTACACTGGCATGCTGATGATGATGATGATGATGATGATGATGATGAT 1734
QY 1820 CCACATCTGGTTGCACTCGTCATGATGGCTCGCTGCATCTCCCTCAGTCCCAAAATCTTA 1879
Db 1735 CCACATCTGGTTGCACTCGTCATGATGGCTCGCTGCATCTCCCTCAGTCCCAAAATCTTA 1794
QY 1880 GTAGCACAAGTTCCTGCGAGAGCTGCTATGTTGCTGCTGGCTGCCCAAGGAGACCTCCTG 1939
Db 1795 G-AGCACAAGTTCCTGCGAGAGCTGCTATGTTGCTGCTGGCTGCCCAAGGAGACCTCCTG 1852
QY 1940 CAGAGCATTTCCTGGTAAAGGACACTTAAAGAGGACATTCATTAAGAGACATTCAGGCAAGATCCTCC 2059
Db 1853 CAGAGCATTTCCTGGTAAAGGACACTTAAAGAGGACATTCATTAAGAGACATTCAGGCAAGATCCTCC 1912
QY 2000 CAGAGCCTTTTGTAGAGCTTCCTGATGTCATTCATTAAGAGACATTCAGGCAAGATCCTCC 2059
Db 1913 CAGAGCCTTTTGTAGAGCTTCCTGATGTCATTCATTAAGAGACATTCAGGCAAGATCCTCC 1971
QY 2060 AACTGCAATATACCAACCTCTCTGAAATTAATTTGCTTAATTTATATTTCTTTCTTT 2119
Db 1972 AACTGCAATATACCAACCTCTCTGAAATTAATTTGCTTAATTTATATTTCTTTCTTT 2031
QY 2120 TTTTCTTAAAGATTCGCTCTGAAATGACATTCATTCATTCATTCATTCATTCATTCATTCAT 2179
Db 2032 TTTTCTTAAAGTAT- - - - -TGCTCTGAAATGACATTCATTCATTCATTCATTCATTCATTCATTCAT 2089
QY 2180 ATTTAGCCTTCCAGTAAATTTTATTAATCTATATATATATATATATATATATATATATATAT 2239
Db 2090 ATTTAGCCTTCCAGTAAATTTTATTAATCTATATATATATATATATATATATATATATATAT 2149
QY 2240 GGAGCTATGATTCATTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2299
Db 2150 GGAGCTATGATTCATTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2209
QY 2300 ATTTTCAATACCAACTTAATTTGCTTGTAAATTAAGTTCAGTGAAGTTCAGTGAAGTTCAGTGAAGT 2359
Db 2210 ATTTTCAATACCAACTTAATTTGCTTGTAAATTAAGTTCAGTGAAGTTCAGTGAAGTTCAGTGAAGT 2268
QY 2360 GAATTTCTTGGTAAATTTATCTTGCATTCGAATCTCATGATTCATATGAAATCGCTTTG 2419
Db 2269 GAATTTCTTGGTAAATTTATCTTGCATTCGAATCTCATGATTCATATGAAATCGCTTTG 2328
QY 2420 ACATATCTTTAGACAGAAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2479
Db 2329 ACATATCTTTAGACAGAAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2387
QY 2480 TTTAGGAGTAGCT 2493
Db 2388 TTTAGGAGTAGGT 2401

RESULT 4
US-10-044-090-720
; Sequence 720, Application US/10044090

Sun Oct 24 15:19:28 2004

```

; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 720
; LENGTH: 2645
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 197362.2
US-10-044-090-720

Query Match      82.7%; Score 2125.2; DB 13; Length 2645;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2207; Conservative 0; Mismatches 8; Indels 7; Gaps 7;

QY 272 CCGGAGCGCGGCTGCGGGCCCGGGTCCGGGCGCGCGGTGACACAGCGCCCTGCTGCAC 331
DB 1 CCGGAGCGCGGCTGCGGGCCCGGGTCCGGGCGCGCGGTGACACAGCGCCCTGCTGCAC 60

QY 332 CAGCTGCGCGCGCGCCCTTCCAGCGGTGCCAGTCTGCTCAGGTGCTCTACTGCGCG 391
DB 61 CAGCTGCGCGCGCGCCCTTCCAGCGGTGCCAGTCTGCTCAGGTGCTCTACTGCGCG 120

QY 392 GCGGCGCGCGCGCGCGCGGACAGCAGGCTTCTGCTGCGCGCGCCCTGATGACCT 451
DB 121 GCGGCGCGCGCGCGCGCGGACAGCAGGCTTCTGCTGCGCGCGCCCTGATGACCT 180

QY 452 GACACCGGCAAGCTGCTCGAGCTGCTGGGCGCTGCCAGAGGACACACGCGCGCAC 511
DB 181 GACACCGGCAAGCTGCTCGAGCTGCTGGGCGCTGCCAGAGGACACACGCGCGCAC 240

QY 512 TTGGGCGAGTTCCAGGCGCGACCGCGCGCGAGTGTGGAGCGCTCTGGAGGTCAA 571
DB 241 TTGGGCGAGTTCCAGGCGCGACCGCGCGCGAGTGTGGAGCGCTCTGGAGGTCAA 300

QY 572 GACGCGCGCGCTGAGGTGGCTGCGCACAGTCTGCTGCGCGAGCCCGCGCTG 631
DB 301 GACGCGCGCGCTGAGGTGGCTGCGCACAGTCTGCTGCGCGAGCCCGCGCTG 360

QY 632 CACCGGTGTGTCAGACTTTCGCCAGTTCGCTGCTTCCCGGACCGGGAAGCGCCCGG 691
DB 361 CACCGGTGTGTCAGACTTTCGCCAGTTCGCTGCTTCCCGGACCGGGAAGCGCCCGG 420

QY 692 GCCGTTTGGAGGAGTGAACCTCTTTATTCCTGAAGCCCGGCGAGTGTGACCTGTC 751
DB 421 GCCGTTTGGAGGAGTGAACCTCTTTATTCCTGAAGCCCGGCGAGTGTGACCTGTC 480

QY 752 GACAGGTGCCAAGACAGTCCAGAAAGAAAGTTCCAGGTGTGTCATCGAAGGACTG 811
DB 481 GACAGGTGCCAAGACAGTCCAGAAAGAAAGTTCCAGGTGTGTCATCGAAGGACTG 540

QY 812 GATCCACGGGTAAACACACCGTGCACCGAGTCAGTGGCAGATTCACTTAAGCGTGTCTC 871
DB 541 GATCCACGGGTAAACACACCGTGCACCGAGTCAGTGGCAGATTCACTTAAGCGTGTCTC 600

QY 872 TTAAGTACACACCTCTTTCATTTGGCGAGTGGAGGAAGATCTTTGATGATGACCAACT 931
DB 601 TTAAGTACACACCTCTTTCATTTGGCGAGTGGAGGAAGATCTTTGATGATGACCAACT 660

QY 932 ATCATTAGAAGAGCTTTTACTCTTTGGGCAATTATATTGGGCTCCGGAATAGCTAAA 991
DB 661 ATCATTAGAAGAGCTTTTACTCTTTGGGCAATTATATTGGGCTCCGGAATAGCTAAA 720

QY 992 GAATCTGCAAAATCTCTCTGTGATTGTAGACAGGTACTGGCAGACGCGCCACCTATGCC 1051
DB 721 GAATCTGCAAAATCTCTCTGTGATTGTAGACAGGTACTGGCAGACGCGCCACCTATGCC 780

```

```

QY 1052 ATAGCCACTGAGTGGGGTCTCCAGCACCTGCCCGCCAGCCCATCCCTGTGTAC 1111
DB 781 ATAGCCACTGAGTGGGGTCTCCAGCACCTGCCCGCCAGCCCATCCCTGTGTAC 840

QY 1112 CAGTGGCCAGAGGACCTGCTCAAACTGACCTTATCTCTGCTGCTGCTGAGTCTGAG 1171
DB 841 CAGTGGCCAGAGGACCTGCTCAAACTGACCTTATCTCTGCTGCTGCTGAGTCTGAG 900

QY 1172 GAGAGTTGTCAGAGGCTGAGGGCGGGGCGGCGATGGAAGACAGGGAAGAGCAGAACTT 1231
DB 901 GAGAGTTGTCAGAGGCTGAGGGCGGGGCGGCGATGGAAGACAGGGAAGAGCAGAACTT 960

QY 1232 GAGGCCAAGAGTGTTCCTCAAAAGGTAGAAATCTCTACACGCGGATGGAGAATCCT 1291
DB 961 GAGGCCAAGAGTGTTCCTCAAAAGGTAGAAATCTCTACACGCGGATGGAGAATCCT 1020

QY 1292 GGTGTCATGTGTTGATGCCAGCCCTCCAGAGAAAGGTCTGCGAGACAGTATTAAAGC 1351
DB 1021 GGTGTCATGTGTTGATGCCAGCCCTCCAGAGAAAGGTCTGCGAGACAGTATTAAAGC 1080

QY 1352 CTAATCCAGAAATGTTTATGTAACCGTAGTTACTCTGCGCAGGTGCCAGTCTAACTAG 1411
DB 1081 CTAATCCAGAAATGTTTATGTAACCGTAGTTACTCTGCGCAGGTGCCAGTCTAACTAG 1140

QY 1412 ATTAGATGTTGTTGAAACATCTACATCCACCATTTGTTATGCAAGTGTCTCCAAATTTCT 1471
DB 1141 ATTAGATGTTGTTGAAACATCTACATCCACCATTTGTTATGCAAGTGTCTCCAAATTTCT 1200

QY 1472 GTTCTCAAGACATGTTGTGTCGAGAAACCTGAGACACAGGATCTTAATTTTACTTTCAG 1531
DB 1201 GTTCTCAAGACATGTTGTGTCGAGAAACCTGAGACACAGGATCTTAATTTTACTTTCAG 1260

QY 1532 CCAATGTCCTCTCTGACTGATGACCGGTATCACAAAGGTCTCTCATCATGCTTC 1591
DB 1261 CCAATGTCCTCTCTGACTGATGACCGGTATCACAAAGGTCTCTCATCATGCTTC 1320

QY 1592 CAGTGAGAGCCAGCGATTCCTTTCTTCTGCGCATAGTAAACATTTCTTGGAAACATATG 1651
DB 1321 CAGTGAGAGCCAGCGATTCCTTTCTTCTGCGCATAGTAAACATTTCTTGGAAACATATG 1380

QY 1652 TTTTCACTTAATCACTACCAATATCTGGAAAGACTGTCTTACTCAGACAGCAGCAGTGT 1711
DB 1381 TTTTCACTTAATCACTACCAATATCTGGAAAGACTGTCTTACTCAGACAGCAGCAGTGT 1440

QY 1712 ACAGAGCAGCAGCAGAGATCTCCAGATCAGAGGAGAGACCCCGAGCGCTCTGCTTCTC 1771
DB 1441 ACAGAGCAGCAGCAGAGATCTCCAGATCAGAGGAGAGACCCCGAGCGCTCTGCTTCTC 1500

QY 1772 CTACACTGGCATGCTGATGAGATCGTGACATGCCCAATTTGGTCTTCCACATCTGTTT 1831
DB 1501 CTACACTGGCATGCTGATGAGATCGTGACATGCCCAATTTGGTCTTCCACATCTGTTT 1560

QY 1832 GCATCTGTCATGATGGGCTCGTGATCTCCCTCAGTCCCAATTTCTAGTAGCAGAGTGT 1891
DB 1561 GCATCTGTCATGATGGGCTCGTGATCTCCCTCAGTCCCAATTTCTAGTAGCAGAGTGT 1619

QY 1892 TCCTGCGAGGCTGCTGATGTCCTGCTGCCAAGAGGACATCTCTGCGAGAGCATTCTT 1951
DB 1620 TCCTGCGAGGCTGCTGATGTCCTGCTGCCAAGAGGACATCTCTGCGAGAGCATTCTT 1678

QY 1952 TGGGTAAAGAACACTTAAAGAGGCAATGATCTTGTGTGAGGCTCAGAGGCGCTTTT 2011
DB 1679 TGGGTAAAGAACACTTAAAGAGGCAATGATCTTGTGTGAGGCTCAGAGGCGCTTTT 1738

QY 2012 GATAGGCTTCTGATGTCATTAAGACATTTCAAGCAAGTCTCCAACTGCAATAT 2071
DB 1739 GATAGGCTTCTGA-GTCATATTAAGACATTTCAAGCAAGTCTCCAACTGCAATAT 1797

QY 2072 ACCAAGCTTCTGAATATATTTTGTCTTATTTATTTCTTTTCTTTTCTTAAAGAA 2131
DB 1798 ACCAAGCTTCTGAATATATTTTGTCTTATTTATTTCTTTTCTTTTCTTAAAGAA 1857

```



```
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: THEREOF
; FILE REFERENCE: CL001305 DIV-II
; CURRENT APPLICATION NUMBER: US/10/681,223
; CURRENT FILING DATE: 2003-10-09
; PRIOR APPLICATION NUMBER: 10/277,032
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/984,980
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 20966
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-681-223-3

Query Match      44.0%; Score 1130.4; DB 16; Length 20966;
Best Local Similarity 98.1%; Pred. No. 1.5e-268; Indels 8; Gaps 8;
Matches 1228; Conservative 0; Mismatches 16;

QY 1243 TGTGTTTCGTCAAAAGGTAGAAATGTCCTACACGCGGATGAGAAATCCTGGCTGCCATGT 1302
DB 17724 TTTGCGTGTCCCGACGGGTAGAAATGTCCTACACGCGGATGAGAAATCCTGGCTGCCATGT 17783
QY 1303 GGTGTATGCGAGCCCTCCAGAGAAAGGTCTTCGACACAGTATTAGCCCTAATCCAGAA 1362
DB 17784 GGTGTATGCGAGCCCTCCAGAGAAAGGTCTTCGACACAGTATTAGCCCTAATCCAGAA 17843
QY 1363 TAGTTTATGTAACCGTAGTTACTCTGCGCAGGTGCCACGCTTAACTAGATTAGATGTTG 1422
DB 17844 TAGTTTATGTAACCGTAGTTACTCTGCGCAGGTGCCACGCTTAACTAGATTAGATGTTG 17903
QY 1423 TTTGAAACATCTACATCCACCAATTTGTTATGCAAGTGTTCCTCAATTTCTGTTCTACAGC 1482
DB 17904 TTTGAAACATCTACATCCACCAATTTGTTATGCAAGTGTTCCTCAATTTCTGTTCTACAGC 17963
QY 1483 ATGTTGTGTGGCAGAAACTGGAGACGAGCATCTTAAATTTTACTTCAGCCATCGTACCC 1542
DB 17964 ATGTTGTGTGGCAGAAACTGGAGACGAGCATCTTAAATTTTACTTCAGCCATCGTACCC 18023
QY 1543 TC-TTTCGACTGATGAGCCGCTCATCAAAAGGTCCCTCTCATCATGTTCCAGTGAGAGG 1601
DB 18024 TC-TTTCGACTGATGAGCCGCTCATCAAAAGGTCCCTCTCATCATGTTCCAGTGAGAGG 18083
QY 1602 CCAGCGATTGCTTTCTTCTGGCATAGTAAACATTTTCTGGACATATGTTTCACTTAA 1661
DB 18084 CCAGCGATTGCTTTCTTCTGGCATAGTAAACATTTTCTGGACATATGTTTCACTTAA 18143
QY 1662 TCACCTACCAATATCTGGAAGACCTGCTTACTCAGACAGCACGAGGTGACAGAAGCAG 1721
DB 18144 TCACCTACCAATATCTGGAAGACCTGCTTACTCAGACAGCACGAGGTGACAGAAGCAG 18203
QY 1722 CAGACAAGATCTTCAGATCAGCAGGAGACCCGAGCCCTCTGCTTCTCTCTACACTGGC 1781
DB 18204 CAGACAAGATCTTCAGATCAGCAGGAGACCCGAGCCCTCTGCTTCTCTCTACACTGGC 18263
QY 1782 ATGCTGATGAGATCGTGACATGCCACATTTGGCTTTCTCCATCTGGTTGCACTCGTCA 1541
DB 18264 ATGCTGATGAGATCGTGACATGCCACATTTGGCTTTCTCCATCTGGTTGCACTCGTCA 18323
QY 1842 TGATGGGTGCTGCTGATCTCCCTCAGTCCCAATTTCTAGTAGCCCAAGTGTTCCTGCAGAG 1901
DB 18324 TGATGGGTGCTGCTGATCTCCCTCAGTCCCAATTTCTAGTAGCCCAAGTGTTCCTGCAGAG 18382
QY 1902 GCTGTCTATGTCTCTGGCTGCCCCAAGGACATCTCCTGCAGAGCCATTTTGGGTAAGGA 1961
DB 18383 GCTGTCTATGTCTCTGGCTGCCCCAAGGACATCTCCTGCAGAGCCATTTTGGGTAAGGA 18441
QY 1962 ACACCTTAAAGAGGCAATGTATCTTTGTCTGAGGCTCAGAGCCCTTTGATAGGCTTC 2021
DB 18442 ACACCTTAAAGAGGCAATGTATCTTTGTCTGAGGCTCAGAGCCCTTTGATAGGCTTC 18501
```

```
QY 2022 TGAATGTCTCATTAAGACATTCAGGCCAAGATGCTCCAACTGCAAATATACCAACCTTC 2081
DB 18502 TGA-GTCATATATAAGACATTCAGGCCAAGATGCTCCAACTGCAAATATACCAACCTTC 18560
QY 2082 TCTGAATATATATTTGCTTATTTATATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 2141
DB 18561 TCTGAATATATATTTGCTTATTTATATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 18619
QY 2142 ATAGATGACATTTTCCATCTGAATGATGCATATCATTTAGCCCAATCCAGTAATTTA 2201
DB 18620 ATAGATGACATTTTCCAT-TGAATCTGGATGCAITTCATTTAGCCCAATCCAGTAATTTA 18678
QY 2202 TTTATATTAAATCTATACATAATATGTTTCTCAGCATAGGAGCTATGATTCATTAATAA 2261
DB 18679 TTTATATTAAATCTATACATAATATGTTTCTCAGCATAGGAGCTATGATTCATTAATAA 18738
QY 2262 AAGTGAGTCAAACCGCTAAATGCAATGTTTGTGTGTTATTTTCATACACAACTTAAT 2321
DB 18739 AAGTGAGTCAAACCGCTAAATGCAATGTTTGTGTGTTATTTTCATACACAACTTAAT 18798
QY 2322 TTGCTCTGTTAAATAAGTTTCAAGTGGATCTTGGAGTGGGATTTCTTGGTAAATTAATCTTG 2381
DB 18799 TTGCTCTGTTAAATAAGTAC-AGTGGATCTTGGAGTGGGATTTCTTGGTAAATTAATCTTG 18857
QY 2382 CACTTGAATGCTCTCATGATTACATATGAAATCGCTTTGACATATCTTTAGACAGAAAAA 2441
DB 18858 CACTTGAATGCTCTCATGATTACATATGAAATCGCTTTGACATATCTTTAGACAGAAAAA 18917
QY 2442 GTAGCTGAGTGAGGGGAAATTTAGAGCTGTGTGACTTTAGGAGTAGCT 2493
DB 18918 GTAGCTGAGTGAGGGGAAATTTAGAGC-TGTGTGACTTTAGGAGTAGCT 18968
```

```
RESULT 7
US-10-264-237-104
; Sequence 104, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 104
; LENGTH: 1579
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-237-104
```

```
Query Match      40.3%; Score 1035.4; DB 16; Length 1579;
Best Local Similarity 98.8%; Pred. No. 1.4e-245;
Matches 1127; Conservative 0; Mismatches 6; Indels 8; Gaps 8;

QY 1353 TAAATCCAGAAATAGTTTGTAGTGAACCGTAGTTACTCTGCCAGGTG-CACGTCTAACATGA 1412
DB 1 TAAATCCAGAAATAGTTTGTAGTGAACCGTAGTTACTCTGCCAGGTG-CACGTCTAACATGA 59
QY 1413 TTAGATGTTGTTTCAAAACATCTACATCCACCAATTTGTTATGACAGTGTCCCAAAATTTCTG 1472
DB 60 TTAGATGTTGTTTCAAAACATCTACATCCACCAATTTGTTATGACAGTGTCCCAAAATTTCTG 119
QY 1473 TTCTACAGCATGTTGTGTGGCAGAAACTGGAGCACAGGATCTTAATTTTACTTTCAGC 1532
DB 120 TTCTACAGCATGTTGTGTGGCAGAAACTGGAGCACAGGATCTTAATTTTACTTTCAGC 179
QY 1533 CATCGTACCTCTTCTGACTGATGGACCCGTCATCACAAAGTCCCTCTCATCATGTGTC 1592
```



```

; NAME/KEY: misc feature
; LOCATION: (261)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (310)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (315)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (347)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (358)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (366)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (367)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (376)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (381)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (386)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (394)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (401)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (402)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (405)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (412)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (413)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (416)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (423)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (432)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (437)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (443)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (448)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature

; LOCATION: (453)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (460)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (464)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (467)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (470)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (479)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (481)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (485)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (487)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (488)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (492)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (493)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-5564

Query Match      8.5%; Score 219.4; DB 9; Length 497;
Best Local Similarity 90.7%; Pred. No. 9e-44;
Matches 255; Conservative 0; Mismatches 23; Indels 3; Gaps 2;

Qy      852 ATTCACTTAAGGCTGCTCTTTAAAGTCACACCCCTCTTCATTGGCCAGTGAGGAAGA 911
Db      2 ATTCCTTAAGGCTGCTCTTTAAAGTCACACCCCTCTTCATTGGCCAGTGAGGAAGA 61

Qy      912 TCTTTGATGATGACCAACTATCATTAGAAGAGCTTTTACTCTTTGGCAATTATATTG 971
Db      62 TCTTTAATGATGAACCAACTATCATTAGAAGAGCTTTTACTCTTTGGCAATTATATTG 121

Qy      972 TGGCTCCGAAATAGCTAAAGAAATCTGCCAAATCTCTGTGATTTGTAGACAGGTACTGGC 1031
Db      122 TGGCTCCGAAATAGCTAAAGAAATCTGCCAAATCTCTGTGATTTGTAGACAGGTACTGGC 181

Qy      1032 ACAGCAGGCGCCTATGCCATAGCCACTG-AGGTGAGTGGGGTCTCCAGCACCTGCC 1090
Db      182 ACAGCAGGCGCCTATGCCATAGCCACTGAGGTGAGTGGGGTCTCCAGCACCTGCC 241

Qy      1091 CCAGCCCATCACCTGTGTA--CCAGTGGCCAGAGGACCTG 1129
Db      242 CCAGCCCATCANCTGTGTNACCAGTGGGCCAGAGGACCTG 282

RESULT 10
US-10-240-425-827/c
; Sequence 827, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
```



```

; NAME/KEY: misc feature
; LOCATION: (232)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (241)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (248)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (309)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (335)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (349)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (350)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (357)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (384)
; OTHER INFORMATION: n equals a,t,g, or c
;
US-09-783-590-11496

Query Match          4.9%; Score 126.2; DB 9; Length 390;
Best Local Similarity 80.5%; Pred. No. 9.6e-21;
Matches 182; Conservative 0; Mismatches 38; Indels 6; Gaps 3;

Qy      1563  TCATCAAAAGGTCCTCTCATCATGTTCCAGTGAGAGCCAGCGATTGCTTTCTTCCTG 1622
Db      10    TCATCAAAAGGTCCTCTNATCATGTTCCAGTAAGAGCCAGCAATTGNTTCTTCCTG 69

Qy      1623  GCATAGTAACATTTTCTTTGGAACATATGTTTCACTTAATCACTACCAAATATCTGGAAG 1682
Db      70    GNATAGTAACATTTTCTTTGGAACATATGTTTAACTTAATCACTACCAAATATCTGGAG 129

Qy      1683  -ACCTGCTTACTCAGA--CAGCACAGGTGTACAGAAG--CAGCAGACAAGATCTTCC 1736
Db      130    AACCTGCTTACTCAGAACAGNACCAGGTGTACAGAAGGCAGCAGACAAGATCTTCC 189

Qy      1737  AGATCAGCAGGAGACCCCGAGCCTTCTGCTCTCTACACTGGCA 1782
Db      190    AGATCAGCAGGAGACCCCGAGGCCTCTGATTCTTCCTNACA 235

RESULT 12
US-10-156-761-7362
; Sequence 7362, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCES: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02

```

```
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      3.1%; Score 79.8; DB 15; Length 9025608;
Best Local Similarity 45.5%; Pred. No. 5.6e-07;
Matches 322; Conservative 0; Mismatches 382; Indels 3; Gaps 1;

QY      1  CGGGGCGGAGGGCGGCTGCTGAGGGCGCCATGGGCTTCGCCCGCGCGGCTCCTTGCGCGG 60
Db      8812441  CGGCGCTGCGCTGCGCGAACCGGAGGACCCACGCTGCGCGTCTCCAGAACCGCGCGT 8812500

QY      61  GCACACTGTGCGGGCCGCTGCTGCGGGCGGCGGGGTGTGCGTGGGGCCATGCTCGGCC 120
Db      8812501  ACCGCACCGCGGTGGTTCGTGCGCCCTGCTGCTGCGCGAGACCTGTGAGGAGGCGCGGGC 8812560

QY      121  GTGCGGCTTGTGCTGGAGCTTCCGAGCTTCACCTGCGCTCACTTCGCCCTTAGCGCGCGA 180
Db      8812561  GSGCGCGCGCGCGTCCGCGTTCGACTACGCGGCGGAGGACACGACGTGACGTGACGTGACGA 8812620

QY      181  CGCCCCCGGCGAGCAGACGCGCCCGACCCCGCGCTTGGCGCGCTGTGGGGCCCCCGGGA 240
Db      8812621  GACCCACCCCGAGGCTTACACGCGGACACGGGAGGAGGTATCCGGCACACCGCGGA 8812680

QY      241  CGCGAGCTACTGCTGTGCTGCGCGCTGCGCGTGAACCGCGAGCGCGGTGCGGGGCCCGGGTCCG 300
Db      8812681  ACACGGCGAGCCCGAGTGGGGGTTTCACTCGGCGCGCGTCCGCGTGCACGCTGCGCTACCG 8812740

QY      301  GGGCGCGCGCTGCACGAGCGCTGCTGACCGCGCGCGCTGCGCGCGCGCGCGCTTCCAGCGGTG 360
Db      8812741  GCTGCGCGCGCTGCACAAACCCCATGAGAGCCGACGCGGCGCACCGGCTCTGCGCAGGG 8812800

QY      361  CCAGCTGTCTAGGCTGCTGCTGCTACTGCGCGGGCGGCGCAGGCCCGCGCGCGCGCACAGG 420
Db      8812801  CGGCAGACTGCTGCTGCGACGACTCCAGTCAGGGGCGCCACACCGGTACGTACCGTCTCGC 8812860

QY      421  CTTCTGCTGCGGACCCCTTGGATGACCTGACACCGCGCAAGCGTGTGCTCGAGCTGCT 480
Db      8812861  GGGCGCTTTCAGCTGCGCGGAGGACCGGTGTCACCGTGTCCCGGAGACAGTGGCGGGCGG 8812920

QY      481  GGGCGCTGCGCAGGAGGACACCGCGCGCTGCGCGGAGTTCAGAGGCGGACCGCGCGGG 540
Db      8812921  CTTGCGCTGCAAGGGAGCGCCCGCGCGCATGTGTGTGCTGCGCGCGCATGTGCGCGCGCGG 8812979

QY      541  CCAGCTGTGGCAGCGCTCTTGGAGGTGCAAGAAGCGGAGGCGGCTGCAAGGTGGGTGCGC 600
Db      8812980  --AGACGGCGCGCGCGTCAAGGTGCGCGTCCCGCGCGGATTTCTGCGCGCGCGTGTGCG 8813037

QY      601  ACAGGTGCTGCGCGTCCCGGAGCGCGCGCTGCGACCGCGTGTGCGAGACTTGTCCAGTTC 660
Db      8813038  GCACCGCGCGCGCGCGCTGCGACCGCGCTCCGCGTGGCGCGCGCGCGCGCGCGCGCGG 8813097

QY      661  CGTGTCTTCCCGGACCGGGAAGCGCGCGCGCGCTTTTGGAGGAGT 707
Db      8813098  CGCGTCTCTGACAGGAGTCAACACCGACACGCTCCCGGATCAAGGAGT 8813144
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 22, 2004, 14:23:02 ; Search time 216 Seconds
(without alignments)
8460.362 Million cell updates/sec

Title: US-10-681-223-1

Perfect score: 2571

Sequence: 1 cggggccgagggcggtcg.....aggagggggttttgcaccc 2571

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2571	100.0	2571	4	US-09-984-880-1
2	2571	100.0	2571	4	US-10-277-032-1
3	1130.4	44.0	20966	4	US-09-984-880-3
4	1130.4	44.0	20966	4	US-10-277-032-3
5	76	3.0	1413	4	US-09-266-965-51
6	76	3.0	53500	4	US-09-266-965-76
7	71.6	2.8	2561	4	US-09-616-289-48
8	70.4	2.7	4411529	3	US-09-103-840A-1
9	68.8	2.7	4403765	3	US-09-103-840A-2
10	68.2	2.7	810	1	US-08-642-255-60
11	67	2.6	756	1	US-08-642-255-50
12	64	2.5	1926	4	US-09-249-588A-2
13	64	2.5	1926	4	US-09-410-399-3
14	64	2.5	2580	3	US-09-050-863-2
15	64	2.5	2580	3	US-09-359-081-2
16	64	2.5	3452	2	US-09-130-114-1
17	64	2.5	8705	4	US-09-647-344A-14
18	64	2.5	9600	3	US-08-910-647-1
19	64	2.5	9600	4	US-09-620-925-1
20	64	2.5	10596	1	US-07-884-811-15
21	64	2.5	10596	1	US-07-885-971-15
22	64	2.5	10596	1	US-08-087-783A-15
23	64	2.5	10596	1	US-08-194-088B-15
24	64	2.5	10596	2	US-08-194-087-15
25	64	2.5	10596	5	PCT-US93-04648-15
26	64	2.5	16080	4	US-09-724-566A-48
27	63.8	2.5	8438	1	US-07-945-283-1

c	28	63.2	2.5	1288	1	US-08-440-856A-9	Sequence 9, Appli
	29	63	2.5	4257	2	US-08-690-473-1	Sequence 1, Appli
	30	63	2.5	4257	3	US-09-259-821A-1	Sequence 1, Appli
	31	63	2.5	4257	3	US-08-843-659-1	Sequence 1, Appli
	32	63	2.5	4257	4	US-08-825-288A-1	Sequence 1, Appli
	33	63	2.5	12001	1	US-08-458-568A-11	Sequence 11, Appli
	34	62.2	2.4	12001	1	US-08-458-568A-11	Sequence 11, Appli
	35	62.2	2.4	30001	1	US-08-125-468-1	Sequence 1, Appli
	36	62.2	2.4	30001	2	US-08-474-933-1	Sequence 1, Appli
	37	61.6	2.4	2127	4	US-09-818-789-75	Sequence 75, Appli
	38	61.6	2.4	2561	4	US-09-616-289-48	Sequence 48, Appli
	39	61.4	2.4	4776	2	US-08-852-401-1	Sequence 1, Appli
	40	61	2.4	43280	2	US-08-804-227C-1	Sequence 1, Appli
	41	60.6	2.4	13613	3	US-09-105-537-3	Sequence 3, Appli
	42	60.6	2.4	38506	3	US-09-320-878-19	Sequence 19, Appli
	43	60.6	2.4	38506	4	US-09-141-908-1	Sequence 1, Appli
	44	60.6	2.4	38506	4	US-09-657-440-19	Sequence 19, Appli
	45	59.6	2.3	4257	2	US-08-690-473-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-984-880-1

; Sequence 1, Application US/09984880

; Patent No. 6489153

; GENERAL INFORMATION:

; APPLICANT: Ming-Hui WEI

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: CLO01305

; CURRENT APPLICATION NUMBER: US/09/984,880

; CURRENT FILING DATE: 2001-10-31

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 2571

; TYPE: DNA

; ORGANISM: HomoSapien

; US-09-984-880-1

Query Match 100.0%; Score 2571; DB 4; Length 2571;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CGGGCCGAGGGCGGCTGCTGAGGCGGCCCATGGCTTCGGCCCGCGGCTCTTGGCGCG 60	
Db	1	CGGGCCGAGGGCGGCGGCTGCTGAGGCGGCCCATGGCTTCGGCCCGCGGCTCTTGGCGCG 60	
QY	61	GCACATGTGCGGGCGGCTGCTCGGCGGGGCTGCGTGGGGCCATGGTCCGCC 120	
Db	61	GCACATGTGCGGGCGGCTGCTCGGCGGGGCTGCGTGGGGCCATGGTCCGCC 120	
QY	121	GTGCGGCTTCGTTCCTGGAGCTTCGACCTGACCTGCTCTTCGCCCTAGGCGCGCA 180	
Db	121	GTGCGGCTTCGTTCCTGGAGCTTCGACCTGACCTGCTCTTCGCCCTAGGCGCGCA 180	
QY	181	CGCCCCCGGCGAGCAGACGCCCCCGCTTGGCGGGCTGTTGGGGCCCCCGGA 240	
Db	181	CGCCCCCGGCGAGCAGACGCCCCCGCTTGGCGGGCTGTTGGGGCCCCCGGA 240	
QY	241	GGCGAGCTACTCGCTGTGCGTCCCGGTGACCCCGAGCGCGGCTCGGGGCCCGGGTCCG 300	
Db	241	GGCGAGCTACTCGCTGTGCGTCCCGGTGACCCCGAGCGCGGCTCGGGGCCCGGGTCCG 300	
QY	301	GGCGGGCGGGCTGCACACCGGCTGCTGCACAGCTGCGCGCGGCCCTTCCAGGGTG 360	
Db	301	GGCGGGCGGGCTGCACACCGGCTGCTGCACAGCTGCGCGCGGCCCTTCCAGGGTG 360	
QY	361	CCAGCTGCTCAGGCTGCTCTGTACTTCCCGGGCGGCCAGGCGGGCGGCACAGCAAGG 420	

Db	361	CGAGCTGCTCAGGCTGCTCTGCTACTGCTGCCCGGGCGCCAGGCGGGCGGCACAGCAAGG	420
Qy	421	CTTCTCTGCTGCGCGACCCCTCGATGACCTGACACCCGGGCAAGCGCTGCTCAGAGTGCT	480
Db	421	CTTCTCTGCTGCGCGACCCCTCGATGACCTGACACCCGGGCAAGCGCTGCTCAGAGTGCT	480
Qy	481	GGGCGGCTCCAGGAGGACACGACCCGCGACTTGGCGAGTTTCAGGCGCGACCCGCGCG	540
Db	481	GGGCGGCTCCAGGAGGACACGACCCGCGACTTGGCGAGTTTCAGGCGCGACCCGCGCG	540
Qy	541	CCAGCTGTGGCAGCGCTCTGGGAGGTGCAAGACGGCAGCGGCTGCGAGTGGGCTGCGC	600
Db	541	CCAGCTGTGGCAGCGCTCTGGGAGGTGCAAGACGGCAGCGGCTGCGAGTGGGCTGCGC	600
Qy	601	ACAGTGTGCGCGCTCCCGAGCCCGCGCTGCAACCGGTGGTGCAGATTCGCCAGTTTC	660
Db	601	ACAGTGTGCGCGCTCCCGAGCCCGCGCTGCAACCGGTGGTGCAGATTCGCCAGTTTC	660
Qy	661	CGTGTGCTTCCCGGACCGGGAAGCCCGCGGCGGTTTTGGAGGAGTGACCTCCTTTAT	720
Db	661	CGTGTGCTTCCCGGACCGGGAAGCCCGCGGCGGTTTTGGAGGAGTGACCTCCTTTAT	720
Qy	721	TCTGAAGCCCGGCGAGTGTCTGACCTGTGCAACAGTGCCTCAAAACAGATCCAGAAAG	780
Db	721	TCTGAAGCCCGGCGAGTGTCTGACCTGTGCAACAGTGCCTCAAAACAGATCCAGAAAG	780
Qy	781	AAAGTTCAGGTTGTTGCATCGAAGGACTGATGCGCAGGGTAAACCAACCGGTGACCCA	840
Db	781	AAAGTTCAGGTTGTTGCATCGAAGGACTGATGCGCAGGGTAAACCAACCGGTGACCCA	840
Qy	841	GTGAGTGGCAGATTCACTTAAGCTGTCTCTTAAAGTCAACACCTCTTGCATTGGCCA	900
Db	841	GTGAGTGGCAGATTCACTTAAGCTGTCTCTTAAAGTCAACACCTCTTGCATTGGCCA	900
Qy	901	GTGAGGAAGATCTTTGATGATGAACCAACTATCATTAGAAGCTTTTTACTCTTTGGG	960
Db	901	GTGAGGAAGATCTTTGATGATGAACCAACTATCATTAGAAGCTTTTTACTCTTTGGG	960
Qy	961	CAATTAATTTGGCTCTCGAATAGCTAAAGATCTGCAATCTGCCAATCTCCTGTGATTTAGA	1020
Db	961	CAATTAATTTGGCTCTCGAATAGCTAAAGATCTGCAATCTGCCAATCTCCTGTGATTTAGA	1020
Qy	1021	CAGGTACTGGCAGCAGCGCCACCTATGCCATGAGCACTGAGTGGGGGCTCTCCA	1080
Db	1021	CAGGTACTGGCAGCAGCGCCACCTATGCCATGAGCACTGAGTGGGGGCTCTCCA	1080
Qy	1081	GCACCTGCCCCAGGCCATCACCCTGTGTACCACTGTCAGTGGCCAGAGGACCTGCTCAAACTGA	1140
Db	1081	GCACCTGCCCCAGGCCATCACCCTGTGTACCACTGTCAGTGGCCAGAGGACCTGCTCAAACTGA	1140
Qy	1141	CCTTATCTGCTGCTCACTGTGAGTCTCTGAGGAGGTTGCGAGGCTGCGAGGCGGGG	1200
Db	1141	CCTTATCTGCTGCTCACTGTGAGTCTCTGAGGAGGTTGCGAGGCTGCGAGGCGGGG	1200
Qy	1201	CATGGAAGAAGACAGGGAAGAAGCAGAACTTGAGGGCAACAGTGTCTTTCGTCAAAGGT	1260
Db	1201	CATGGAAGAAGACAGGGAAGAAGCAGAACTTGAGGGCAACAGTGTCTTTCGTCAAAGGT	1260
Qy	1261	AGAAATGTCTTACAGCGATGAGAAATCCTGCTGCCATGTTGATGCCAGCCCTC	1320
Db	1261	AGAAATGTCTTACAGCGATGAGAAATCCTGCTGCCATGTTGATGCCAGCCCTC	1320
Qy	1321	CAGAGAAAGGCTCTGACAGAGTATTAAGCTTAATCCAGATAGTTTGTAGTGAACCGTA	1380
Db	1321	CAGAGAAAGGCTCTGACAGAGTATTAAGCTTAATCCAGATAGTTTGTAGTGAACCGTA	1380
Qy	1381	GTTACTCTGCGCAGGTGCCACGCTAACTAGATTAGATTGTTTGAACATCTCATCC	1440
Db	1381	GTTACTCTGCGCAGGTGCCACGCTAACTAGATTAGATTGTTTGAACATCTCATCC	1440
Qy	1441	ACCATTTGTTATGAGTGTCCAAATTTCTGTTCTACAGCATGTTGTGCGCAAAA	1500
Db	1441	ACCATTTGTTATGAGTGTCCAAATTTCTGTTCTACAGCATGTTGTGCGCAAAA	1500
Qy	1501	CTGGAGACGAGCATCTTAATTTTACTTCAGCCATCGTACCCTCTTCTGACTGATGGAC	1560
Db	1501	CTGGAGACGAGCATCTTAATTTTACTTCAGCCATCGTACCCTCTTCTGACTGATGGAC	1560
Qy	1561	CCTCATCAAAAGTCCCTCTCATCATGTTCCAGTGAGAGGCGAGCATGTCTTTCTTCC	1620
Db	1561	CCTCATCAAAAGTCCCTCTCATCATGTTCCAGTGAGAGGCGAGCATGTCTTTCTTCC	1620
Qy	1621	TGGCATAGTAAACATTTTCTTGAAACATATGTTTCACTTAATCACTACCAAAATCTGGA	1680
Db	1621	TGGCATAGTAAACATTTTCTTGAAACATATGTTTCACTTAATCACTACCAAAATCTGGA	1680
Qy	1681	AGACCTGTCTTACTCAGACGACGAGGTGACAGAGCAGACAGATCTTCAGAT	1740
Db	1681	AGACCTGTCTTACTCAGACGACGAGGTGACAGAGCAGACAGATCTTCAGAT	1740
Qy	1741	CAGCAGGAGAGACCCCGGAGCTCTCTCTCTACACTGGCATGCTGATGAGATCGTAC	1800
Db	1741	CAGCAGGAGAGACCCCGGAGCTCTCTCTCTACACTGGCATGCTGATGAGATCGTAC	1800
Qy	1801	ATGCCACATGCGCTCTTCCACATCTGTTGCACTCTGTCATGAGTGGTCTGTCATCT	1860
Db	1801	ATGCCACATGCGCTCTTCCACATCTGTTGCACTCTGTCATGAGTGGTCTGTCATCT	1860
Qy	1861	CCCTCAGTCCCAAAATTTCTAGTAGCCAGGTTCCTGCGAGGCTGTCTATGTCTCTGGC	1920
Db	1861	CCCTCAGTCCCAAAATTTCTAGTAGCCAGGTTCCTGCGAGGCTGTCTATGTCTCTGGC	1920
Qy	1921	TGCCCAAGGGAACACTCTCTGAGAGCCATTTTGGGTAAAGAACACTTACAAAGAGGCAT	1980
Db	1921	TGCCCAAGGGAACACTCTCTGAGAGCCATTTTGGGTAAAGAACACTTACAAAGAGGCAT	1980
Qy	1981	TGATCTTGTCTGAGGCTCAGAGCCCTTTGATAGGCTTCTGATGCTATTATAAAGAC	2040
Db	1981	TGATCTTGTCTGAGGCTCAGAGCCCTTTGATAGGCTTCTGATGCTATTATAAAGAC	2040
Qy	2041	ATTGAGCCAGATGCTCCAACTGCAATATACCAACCTTCTGTAATTTATATTTTGCCT	2100
Db	2041	ATTGAGCCAGATGCTCCAACTGCAATATACCAACCTTCTGTAATTTATATTTTGCCT	2100
Qy	2101	ATTATATTTCTTTCTTTTCTTAAAGAAATGGCTCTGAATAGAAATGCACATTTTCCA	2160
Db	2101	ATTATATTTCTTTCTTTTCTTAAAGAAATGGCTCTGAATAGAAATGCACATTTTCCA	2160
Qy	2161	CTGAACTGGATGCTATCATTTAGCCATTCAGTAAATTTATTTATTTATTTATTTATAT	2220
Db	2161	CTGAACTGGATGCTATCATTTAGCCATTCAGTAAATTTATTTATTTATTTATTTATAT	2220
Qy	2221	AATATGTTTCTCAGCATAGGAGCTATGATTCATTAATTTAAAGTGGAGTCAAAACGCTA	2280
Db	2221	AATATGTTTCTCAGCATAGGAGCTATGATTCATTAATTTAAAGTGGAGTCAAAACGCTA	2280
Qy	2281	AATGCAATGTTTGTGTTTCTTATTTTCAATACAAACTTAAATTTGCTTGTAAATAGTT	2340
Db	2281	AATGCAATGTTTGTGTTTCTTATTTTCAATACAAACTTAAATTTGCTTGTAAATAGTT	2340
Qy	2341	CAGTGGATCTTGGAGTGGATTTCTTGTGTAATTTCTTGCATTTGAATGCTCATGAT	2400
Db	2341	CAGTGGATCTTGGAGTGGATTTCTTGTGTAATTTCTTGCATTTGAATGCTCATGAT	2400
Qy	2401	TACATATGAATCGCTTTTCACATATCTTTTAGACAGAAAAAGTAGTGTAGTGGGGGAA	2460
Db	2401	TACATATGAATCGCTTTTCACATATCTTTTAGACAGAAAAAGTAGTGTAGTGGGGGAA	2460
Qy	2461	ATTATAGACTTGTGACTTTAGGAGTAGTGTCTTTATACATATCTCAAGCCCTG	2520
Db	2461	ATTATAGACTTGTGACTTTAGGAGTAGTGTCTTTATACATATCTCAAGCCCTG	2520
Qy	2521	AAGCCTTGTATGCTGCGAGCTCGCACTAAAGGAGGGGCTTTTGCACCC	2571
Db	2521	AAGCCTTGTATGCTGCGAGCTCGCACTAAAGGAGGGGCTTTTGCACCC	2571


```

; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 51
; LENGTH: 1413
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-51

Query Match          3.0%; Score 76; DB 4; Length 1413;
Best Local Similarity 50.8%; Pred. No. 2.4e-08;
Matches 181; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 235 CCGGAGCGAGCTACTCGCTGTGCTGCGTGGAGACCCCGGACGCGGCTGCGGGCCGG 294
DB 192 CCGGACCGGATCTACGGCTTACCGACCTCGCCCCCGGCGCGGTACCGCATCACCTT 251
QY 295 GGTTCGGGGCGGCGCGCTCACACAGCGCTGTGACACAGCTGCGCGCGGCCCTTCCA 354
DB 252 CCCCGGTGTGGCAGCTGCGGGAACCCCGCTGTGGGAATCGCCGCCGCGCGCGCG 311
QY 355 GCGGTCCAGCTGCTCAGGCTGTCTGTACTGCTCCCGGGCGGCGAGGCGGCGGCGACA 414
DB 312 CAGGACCGTGTGCTGAACGTGCGGGCACCTACCCCGCCCGCCCATCGAGCGGCTGT 371
QY 415 GCAGAGCTTCTGCTGGCGGACCCCTGTGATGACCTGACACCCGCGAGCGCTGTGCGA 474
DB 372 GGTCTCCGGTTCGTGCGGCCGGAATCGGAGCGCGCGGTGAGCCCGACGCTGTGCC 431
QY 475 GCTGCTGGGCGCTGCGAGGAGCACACGCGCCGCACTTGGGCGAGTTGAGGCGCGACCC 534
DB 432 GCTGCTGGCGGCTCGACTACGAATCGAGCTCGAGGTCGGCGAGCTGCGCGCGACCC 491
QY 535 GCGCGGCGACTGTGCGAGCGCTCTGGGAGGTGCAAGAGCGGAGCGCGGCTGCAGG 590
DB 492 GCGCGCTTCTCGGGCGGCGCTCGGGGCCCTGCGCGCGCCGACCGCGCGGATGG 547

RESULT 6
US-09-266-965-76
; Sequence 76, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 76
; LENGTH: 53500
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-76

Query Match          3.0%; Score 76; DB 4; Length 53500;
Best Local Similarity 50.8%; Pred. No. 1.6e-07;
Matches 181; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 235 CCGGAGCGAGCTACTCGTGTGCTGCGTGGAGACCCCGGACGCGGCTGCGGGCCGG 294

```



```
QY 215 CTGGCGGCGCTGTTGGGGCCCGGAGCGAGCTACTCGCTGTGCTGCTGACCCCG 274
Db 1025 CAGGCT-----CAGGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 980
QY 275 GACCGCGGCTGGGGCCCGCGGTCCCGGGCGCGCGCTGACACAGCGCTGTGTGACACG 334
Db 979 TCGGGCGGCGTGTG-GTGTGCGCGCGCGGTGGCGGGGGGTGTGTGTGTGTGCTG 921
QY 335 CTGGCGCGCGCGCGCTTCCAGCGGTGCCAGCTGCTCAGCGTGTGTGTGTGTGTGTGT 394
Db 920 TGGCGCGCGCGCGCGCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 861
QY 395 GGCCAGCGCGCGCGCGCACAGCAAGCTTCTGCTGCGCGACCGCTGTGATGACCTGAC 454
Db 860 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 801
QY 455 ACCCGGCAAGCGCTGTCTGAGTGTCTGGCGCGCTGCGCAGGAGCACACCGCGCGCTG 514
Db 800 CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 741
QY 515 GCGGAGTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 574
Db 740 GCGGCGGAGGCGCGCGCTGTGCGG-CAGCGCGCGCGCGCGCGCGCGCGCGCGCG 682
QY 575 GCGAGCGCGCTCAGGTGGGCTGCGCACAGGTGTGCGCG 614
Db 681 CCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642

RESULT 8
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 2.7%; Score 70.4; DB 3; Length 4411529;
Best Local Similarity 47.7%; Pred. No. 3.5e-05;
Matches 267; Conservative 0; Mismatches 291; Indels 2; Gaps 2;

QY 91 CGGGTCTCGCTGGGGCCATGCTCCGCGGTGCGCGCTTGTCTGTGAGCTTCCGACTG 150
Db 3944636 CGGCCCCACCGTGGCGCGCGCTTGCATTCCTCCGCGCGCGCTTGC 3944577
QY 151 CACCTTGGCTCACTTCGCGCTAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 210
Db 3944576 CATTGCGGCGCGACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3944517
QY 211 CGCGTGGCGCGCTGTTGGGGCCCGCGGAGCGCAGTACTGCTGTGCGTGTGCGGTGAC 270
Db 3944516 GGGGTCGCGGAGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3944457
QY 271 CCGGAGCGCGGTGCGGGCGCGCGGTTCGGGGCGCGCGGTGCACACAGCGCTGTGCA 330
Db 3944456 CCGCGCGGTGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3944398
QY 331 CCAGTGTGCGCGCGCGCGCTTCCAGCGGTGCCAGTGTCTCAGGCTGCTGTGCTACTGCCC 390
```

```
Db 3944397 CCGGCGCGCGCGCGCGCTTGCAGCGCGCTTGCAGCGCGCTTGCAGCGCGCTTGCAG 3944338
QY 391 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 450
Db 3944337 TCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3944279
QY 451 TGACACCGCGCGCGCGCTGCTGAGGTGCTGAGGTGCTGAGGTGCTGAGGTGCTGAGGT 510
Db 3944278 GCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3944219
QY 511 CTGGCGCGAGTTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570
Db 3944218 GATGGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3944159
QY 571 AGACGCGCGCGCGCTGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630
Db 3944158 AGCGGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3944099
QY 631 GCACCGGTGTGTCAGACT 650
Db 3944098 GGCACCGTCAAGCGCGCGCGCT 3944079

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 2.7%; Score 68.8; DB 3; Length 4403765;
Best Local Similarity 47.5%; Pred. No. 8.5e-05;
Matches 266; Conservative 0; Mismatches 292; Indels 2; Gaps 2;

QY 91 CGGGTCTCGCTGGGGCCATGCTCCGCGGTGCGCGCTTGTCTGTGAGCTTCCGACTG 150
Db 3938183 CGGCCCCACCGTGGCGCGCGCTTGCATTCCTCCGCGCGCGCTTGC 3938124
QY 151 CACCTTGGCTCACTTCGCGCTAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 210
Db 3938123 CATTGCGGCGCGACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3938064
QY 211 CGCGTGGCGCGCTGTTGGGGCCCGCGGAGCGCAGTACTGCTGTGCGTGTGCGGTGAC 270
Db 3938063 GGGGTCGCGGAGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3938004
QY 271 CCGGAGCGCGCTGCGGGCGCGCGGTTCGGGGCGCGCGGTGCACACAGCGCTGTGCA 330
Db 3938003 CCGCGCGGTGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3937945
QY 331 CCAGTGTGCGCGCGCGCGCTTCCAGCGGTGCCAGTGTCTCAGGCTGCTGTGCTACTGCCC 390
Db 3937944 CCGGCGCGCGCGCGCGCTTCCAGCGGTGCCAGCGGTGCCAGCGGTGCCAGCG 3937885
```



```
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLSCULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-50

Query Match      2.5%; Score 67; DB 1; Length 756;
Best Local Similarity 45.3%; Pred. No. 2.4e-06;
Matches 287; Conservative 0; Mismatches 340; Indels 6; Gaps 1;

QY 76 GCTGCTCGGGCGGGGCTGCGCTGGGCGCATGGCTCCGCGTGCCTTGGTCTCT 135
DB 1 GGTGCTCAGCGCCACAGCAGGTTCGAAGGGCGCGCATGGCCACAGAGCCCGAAAGGTGCG 60
QY 136 GGAGCTTCCCGACTGCACCTTGGCTCACTTTCGCTTAGGCGCCGAGCGCCCCCGGCGACGC 195
DB 61 CTGGACCGGCTGGTCCACCGGCTGCTCCGGACCTGCAGGCGCGCCAGGTGCGCTGGA 120
QY 196 AAGAGCCCCCGACCCCGCTGGCGGCTGTGGGGCCCGCGAGCGGCGAGCTACTCGCT 255
DB 121 CCGGCTGGTCCACCGGCTGCTCCGGACCTGCAGGCGCCCGCAGGTGCGCTGACCGGCT 180
QY 256 GTGCGTGGCCGCTGACCCCGGAGCGCGGCTGCGGGGCCCGGGTCCGGGCGGCGGTGCA 315
DB 181 GGTACACCGGCTGCTCCGGGACTGTCAGAGCCCGCAGGTGCGCTGAGACCGGCTGTCCA 240
QY 316 CCAGCGCTGTGTCACACAGCTGCGCGCGGCCCGCTTTCACGCGTGCACGCTCAAGCT 375
DB 241 CCGGCTGTGTCAGGACCTGCAGGCGCCCGCAGGTGCGCTTGGACCGGCTGTCCACCGGT 300
QY 376 GCTCTGCTACTGCCCGGGGCGCAGCGCGGCGCGGCGCAGCAAGGCTTCTGCTGGCGCA 435
DB 301 GCTCCGGGACCTGCAGGCGCGCGCAGGTGCGCTGAGCCGCGTGTCCACCGGCTGCTCG 360
QY 436 CCCCCTGGATGACCTGTACACCGGCAAGCGTGTGCTGAGCTGCTGGGCGCCTGCCAGGA 495
DB 361 GGACCTGACGGCCCGCAGAGTGGGCTGTGACCGGCTGTCCACCGGCTGCTCCGGAGCT 420
QY 496 GGACACACCCCGCACTTGGGCGAGTTCAGAGCCGACCCCGCGCGGCGCAGCTGTGCGAGCG 555
DB 421 GCAGGCGCCCGCAGGTGCGCTGCAGCGGCTGGTCCACCGGCTGCTCCGGGACCTGCAGGC 480
QY 556 CTTCTGGGAGGTGCAAGACGGCAGGCGGCTGCAGGTGGGCTGGCGACAGGTCTGTGCC-- 613
DB 481 CCGCCAGGTGCGCTGAGCCGCGTGTGTCACCGGCTGCTCCGGGACCTGCAGGCGCGCGCA 540
QY 614 ----GTCCCGGAGCCCGCTGCACCGGCTGGTGCCAGACTTGCCCGAGTTCCGTTGTTT 669
DB 541 GGTGCGCTTGGACCGGCTGTGTCACCGGCTGTCCCGGACCTGCAGGCGCGCGAGGTGCG 600
QY 670 CCGGACCGGGAAGCGCCCGGCGCGTTTGG 702
DB 601 CTGGACCGGCTGGTCCACCGGCTGCTCCGGGA 633

RESULT 12
US-09-249-585A-2/c
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISODES
; FILE REFERENCE: 0867/0905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

Query Match      2.5%; Score 64; DB 4; Length 1926;
Best Local Similarity 46.3%; Pred. No. 2.1e-05;
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 25 GCGCGCCATGGGCTTTCGCGCGCGGCTCTGCGCGGCGCACATGTGCGGGCGCGCTGCTCG 84
DB 719 GCGGCTCTGCTGCTGCGCGCTCTGCTGCTGCGCGCGCTCTGCTGCGCGCTCTGCTGCG 660
QY 85 GCGGCGCGGGGTCTGGGCTGGGGCCATGGCTCGCGCGTCCCGTTCGCTGCTGAGGCTTCC 144
DB 659 GGTCTGCGCCCTCTGCTGCTGCGCGCTCTGCGCGCTCTGCGCGCTCTGCTGCTGCGCC 600
QY 145 CGACTGCACCCCTGGCTCACTTCGCTAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 204
DB 599 CTTCTGCTCTGCGCGCTCTGCGCGCTCTGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTG 540
QY 205 CGACCCCGCGCTGGGCGGCTGTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 264
DB 539 CTTCTGCTGCTGCTGCGCGCTCTGCTGCTGCTGCTGCGCGCTCTGCTGCTGCTGCTGCT 480
QY 265 CTTGACCCCGGAGCGGCTGCGGGCGCGGCTGCGGGCGCGGCTGCGGGCGCGGCTGCGAG 324
DB 479 CTTGCTGCTGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 325 GGTGCACACAGTGCAGCGCGCGCGCTTCCAGCGGTGCGAGCTGCTGCTGCTGCTGCTA 384
DB 419 CTTGCGCTCTGCTGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 385 CTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 444
DB 359 CTTCTGCGCGCTCTGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 445 TGACCTGACACCGCGGCAAGCGCTGCTGAGTGTCT 480
DB 299 GCTCTGCGCGCTCTGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 264

RESULT 13
US-09-410-399-3/c
; Sequence 3, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; FILE REFERENCE: To Genomic Host DNA
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein-Barr virus
US-09-410-399-3

Query Match      2.5%; Score 64; DB 4; Length 1926;
Best Local Similarity 46.3%; Pred. No. 2.1e-05;
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 25 GCGCGCCATGGGCTTTCGCGCGCGGCTCTGCGCGGCGCACATGTGCGGGCGCGCTGCTCG 84
```



```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

Query Match      2.5%; Score 64; DB 3; Length 2580;
Best Local Similarity 46.3%; Pred. No. 2.4e-05;
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 25 GGCGCCCATGGCTTCGCCCGCGGCTCCTCGCGGGCCACATGTGCGGGCGCGCTGCTCGG B4
Db 1102 GCGCCTCCTGCTCTGCGCCCTCTGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCT 1043
QY 85 GCGCGCGGGGTCTGCGCTGCGGCCATGGCTCGCGCGTTCGTCCTGAGCTTCC 144
Db 1042 GCTCTGCGCCCTCCTGCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCTCTGCGCCCT 983
QY 145 CGACTGCACCCCTGGCTCACTTGCGCCCTAGCGCGGAGCGCCCGCGGACGACGCCCC 204
Db 982 CTTCTGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCC 923
QY 205 CGACCCCGCGCTGCGCGGCTGTGGGGCGCCCGGAGCGAGCTACTCGCTGTGCGTGCC 264
Db 922 CTTCTCTGCTCTGCGCCCTCTGCTCTGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCT 863
QY 265 CGTGACCCCGGAGCGCGGCTGCGGGGCCCGGGTTCGGGGCGGGCGGCTGCACCGCGCT 324
Db 862 CCGTCTCTGCGCCCTCTGCGCCCTCTGCTCTGCTCTGCGCCCTCTGCTCTGCGCCCTCT 803
QY 325 GCTGCACACAGCTGCGCGGCGCCCTTCCAGCGGTGCCAGCTGCTCAGGCTGCTGCTA 384
Db 802 CTTGCGCCCTCTGCGCCCTCTGCTGCTCTGCGCCCTCTGCTGCTCTGCGCCCTCTGCGCC 743
QY 385 CTGCGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 444
Db 742 CTTCTGCGCCCTCTGCGCCCTCTGCTGCTCTGCGCCCTCTGCTGCTCTGCGCCCTCTGCT 683
QY 445 TGACCTTGACACCGGCGGAGCGGTGCTCGAGCTGCT 480
Db 682 GCTCTGCGCCCTCTGCGCCCTCTGCTGCTCTGCTGCTCTGCTGCTCTGCTGCTCTGCT 647
```

Search completed: October 22, 2004, 20:24:32
Job time : 235 secs

This Page Blank (uspto)

CC to the kinase, an isolated nucleic acid molecule encoding the kinase
CC proteins detailed above (appearing as the cDNA, AD63369, and gene,
CC ADC63371), a gene chip comprising the nucleic acids, a transgenic non-
CC human animal comprising the nucleic acids; a nucleic acid vector
CC comprising the nucleic acids, a host cell containing the vector,
CC detecting the presence of the kinase in a sample, detecting the presence
CC of the nucleic acids in a sample, a pharmaceutical composition comprising
CC a kinase binding agent, treating a disease or condition mediated by a
CC human kinase protein and identifying a modulator of the expression of the
CC kinase involves. The kinase is useful for identifying a modulator, for
CC identifying an agent that binds to the kinase, for biological assays
CC related to kinases, drug screening assays, identifying compounds that
CC modulate kinase activity, to screen a compound for the ability to
CC stimulate or inhibit interaction between the kinase protein and a
CC molecule that normally interacts with kinase protein, for competitive
CC binding assays to discover compounds that interact with kinase,
CC pharmacogenomic analysis and treating a disorder characterised by an
CC absence of inappropriate or unwanted expression of the protein, useful as
CC target spot diagnosing active protein activity, disease or predisposition
CC to disease. The nucleic acid is useful for assessing expression in
CC disease states, diagnostic tools as an immunological marker for aberrant
CC protein for inhibiting protein function, tissue typing. The nucleic acids
CC are useful for facilitating protein trafficking or facilitate
CC manipulation of a protein for assay or production, probes, primers,
CC chemical intermediates, constructing recombinant vectors, expressing
CC antigenic portions of the proteins, for determining chromosomal positions
CC of the nucleic acid molecules, for making vectors containing the gene
CC regulatory regions, as hybridisation probes for determining the presence,
CC level, form and distribution of nucleic acid expression. The gene
CC encoding the kinase is located on human chromosome 2. The present
CC sequence is the cDNA encoding the kinase.

XX
SQ Sequence 2571 BP; 561 A; 711 C; 675 G; 624 T; 0 U; 0 Other;

Query Match 100.0%; Score 2571; DB 10; Length 2571;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGGCGAGGCGGGTGGCTGAGGCGCCGATGGCTTCGCGCGCGGCTCTGGGCGG 60
DB |||||
QY 1 CGGGGCGAGGCGGGTGGCTGAGGCGCCGATGGCTTCGCGCGCGGCTCTGGGCGG 60
DB |||||
QY 61 GCCACTGTGCGGCGCGCTCTCGGGCGGCGGGGTCGCGTGGGGCCATGGCTCCGCC 120
DB |||||
QY 61 GCCACTGTGCGGCGCGCTCTCGGGCGGCGGGGTCGCGTGGGGCCATGGCTCCGCC 120
DB |||||
QY 121 GTGCGGCTTCGTCTGGAGCTTCGCACTGCCACCTGGGTCACTTCGCCCTAGGCGCGA 180
DB |||||
QY 121 GTGCGGCTTCGTCTGGAGCTTCGCACTGCCACCTGGGTCACTTCGCCCTAGGCGCGA 180
DB |||||
QY 181 CCGCCCGGCGAGCGAGCGCCCGACCCCGCTGGCGGCGCTGTGGGGCCCCCGGA 240
DB |||||
QY 181 CCGCCCGGCGAGCGAGCGCCCGACCCCGCTGGCGGCGCTGTGGGGCCCCCGGA 240
DB |||||
QY 241 GCGGAGTACTGCTGTGGTGGCTGACCCCGGACCGCGTGGGGCGCCGGGTCCG 300
DB |||||
QY 241 GCGGAGTACTGCTGTGGTGGCTGACCCCGGACCGCGTGGGGCGCCGGGTCCG 300
DB |||||
QY 301 GCGGCGCGGCTGCAACGAGCGCTGTGACCCCGGCGCGGCGCCCTTCAGCGGTG 360
DB |||||
QY 301 GCGGCGCGGCTGCAACGAGCGCTGTGACCCCGGCGCGGCGCCCTTCAGCGGTG 360
DB |||||
QY 361 CCAGCTGCTCAGGCTGCTCTGCTACTGCGCGGCGCGAGCGCGGCGGCGACGCMAGG 420
DB |||||
QY 361 CCAGCTGCTCAGGCTGCTCTGCTACTGCGCGGCGCGAGCGCGGCGGCGACGCMAGG 420
DB |||||
QY 421 CTTCTGCTGCGCGACCCCTCGATGACCCCTGACACCCCGGCAAGCGCTGCTCGAGTGT 480
DB |||||
QY 421 CTTCTGCTGCGCGACCCCTCGATGACCCCTGACACCCCGGCAAGCGCTGCTCGAGTGT 480
DB |||||
QY 481 GGGGCGCTCCAGGAGGCAACAGCCCGCACTTGGGCGAGTTCGAGGCGCGA CCGCGCGG 540
DB |||||
QY 481 GGGGCGCTCCAGGAGGCAACAGCCCGCACTTGGGCGAGTTCGAGGCGCGA CCGCGCGG 540
DB |||||

[illegible]

RESULT 2

RESULT 2
ABX93989

ABX93989
IN ABY03000 standard: CDNA: 2571 BB

ID
yy
ABX93989

XX

AC ABX93989;

XX

Human; thymidylate kinase subfamily; kinase; therapeutic agent; immune response; kinase associated disorder; tissue typing; pharmacogenomic analysis; gene; ss.

Homo sapiens.

Key	Location/Qualifiers
5'UTR	1..31
	/*tag= a
CDS	32..1381
	/*tag= b
	/product= "Kinase"
3'UTR	1382
	/*tag= c

US6489153-B1

03-DEC-2002.

31-OCT-2001: 2001US-00984880

31-OCT-2001: 2001US-00984880

(PEKE) PE CORP NY

Wei M. Ketchum KA. Beasley EM. DiFrancesco V:

WPT: 2003-327246/31.

P-PSDB; AB008678.

New human kinase pot
development of human

nucleic acid molecule disorder associated

Claim 5: Fig 1A-C: 49pp: English.

The invention describes a new isolated nucleic acid molecule encoding a kinase of the thymidylate kinase subfamily. The human kinase polypeptides and nucleic acid molecules are useful as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, or serve as targets for the development of human therapeutic agents that modulate kinase activity in cells and tissues that express the kinase. The kinase proteins are also useful in raising antibodies or eliciting another immune response, as reagents in assays designed to quantitatively determine levels of the protein in biological fluids, as markers for tissues in which the corresponding protein is preferentially expressed, or in screening a compound for the ability to stimulate or inhibit interaction between the kinase protein and a molecule that normally interacts with the kinase protein. The kinase proteins and nucleic acid molecules can also be used in providing a target for diagnosing a disease or a predisposition to a disease mediated by the polypeptide or nucleic acid molecule, or for treating a disorder associated with the absence of, inappropriate, or unwanted expression of the protein or nucleic acid molecule. The antibodies are useful for tissue typing, in isolating or detecting kinase proteins, or in pharmacogenomic analysis. This sequence encodes the novel human thymidylate kinase subfamily kinase

Sequence 2571 BP: 561 A: 711 C: 675 G: 624 T: 0 U: 0 Other:

Very Match 100.0%: Score 2571: DB 10: Length 2571;

Best Local Similarity 100.0%; F1ed: NS: 0
Matches 2571: Conservative 0: Mismatches

1 CCGGGCCGAGGGCGGGCTGGCTGAGGGCGCCCATGGCCCTTCGCCCGCGCGCTCCTGCGCG 60

60

[illegible]

DT 10-JUN-2003 (first entry)

Db	61	GCCACTGTGCGGGCCGCTGCTCGGGGGCGCGGGTCTGGGCTGGGGCCATGGCTCCGCC	120
QY	121	GTGCGGCTTGTCTTGGAAGTTCCTCGACTGCACTTGGGTCACTTGGCCCTTAGGCGCCGA	180
Db	121	GTGCGGCTTGTCTTGGAAGTTCCTCGACTGCACTTGGGTCACTTGGCCCTTAGGCGCCGA	180
QY	181	CGCCCCGGGCGAGCAGACGCCCCCGACCCCGCTGCGGGCGCTGTGCGGCCCCCGGA	240
Db	181	CGCCCCGGGCGAGCAGACGCCCCCGACCCCGCTGCGGGCGCTGTGCGGCCCCCGGA	240
QY	241	GCGAGCTACTGGCTGTGGCTGCCCTGACCCCGGACCGCGGTGCGGGGCGCCGGGTCCG	300
Db	241	GCGAGCTACTGGCTGTGGCTGCCCTGACCCCGGACCGCGGTGCGGGGCGCCGGGTCCG	300
QY	301	GGCGGCGGGCTGCACACAGGCGTGTGTGCAACAGCTGCGCGCGGGCCCTTCCAGCGGTG	360
Db	301	GGCGGCGGGCTGCACACAGGCGTGTGTGCAACAGCTGCGCGCGGGCCCTTCCAGCGGTG	360
QY	361	CCAGCTGTCTCAGGCTGCTGTGCTACTGCGCGGGCGCCAGGCGGGCGGCACAGCAAG	420
Db	361	CCAGCTGTCTCAGGCTGCTGTGCTACTGCGCGGGCGCCAGGCGGGCGGCACAGCAAG	420
QY	421	CTTCTGTCTGCGCGACCCCTCGATGACCTTGACACCCCGGCAAGCGCTGCTCGAGTGTCT	480
Db	421	CTTCTGTCTGCGCGACCCCTCGATGACCTTGACACCCCGGCAAGCGCTGCTCGAGTGTCT	480
QY	481	GGGCGCTCCGAGGAGGCAACGCGCCGCACTTGGGCGAGTTGAGGCGCGACCCGCGGG	540
Db	481	GGGCGCTCCGAGGAGGCAACGCGCCGCACTTGGGCGAGTTGAGGCGCGACCCGCGGG	540
QY	541	CCAGCTGTGCGACGCGCTCTGGGAGGTGCAAGACGCGAGCGCTGCGAGTGGCTGGCG	600
Db	541	CCAGCTGTGCGACGCGCTCTGGGAGGTGCAAGACGCGAGCGCTGCGAGTGGCTGGCG	600
QY	601	ACAGTGTGCGCGCTCCCGAGCCCCGCTGCAACCGGTTGGTGCCAGACTTGCCAGTTTC	660
Db	601	ACAGTGTGCGCGCTCCCGAGCCCCGCTGCAACCGGTTGGTGCCAGACTTGCCAGTTTC	660
QY	661	CGTGTCTTCCGGACCGGAGCGCGCGGGCCGTTTGGAGAGTGTACCTCTTTAT	720
Db	661	CGTGTCTTCCGGACCGGAGCGCGCGGGCCGTTTGGAGAGTGTACCTCTTTAT	720
QY	721	TCTGAAGCCCGGGAGTCTTGAACCTGTGCAACAGTGCACCAAGTGCACCAAGATCCAGAAAG	780
Db	721	TCTGAAGCCCGGGAGTCTTGAACCTGTGCAACAGTGCACCAAGTGCACCAAGATCCAGAAAG	780
QY	781	AAAGTTCAGGTTGTGCCATCGAAGGACTGGATGCCACGGGTAAACACACGCTGACCCA	840
Db	781	AAAGTTCAGGTTGTGCCATCGAAGGACTGGATGCCACGGGTAAACACACGCTGACCCA	840
QY	841	GTCAGTGGCAGATTCACTTAAGGCTGTCTCTTAAAGTCAACACCTCTTGCATTGGCCA	900
Db	841	GTCAGTGGCAGATTCACTTAAGGCTGTCTCTTAAAGTCAACACCTCTTGCATTGGCCA	900
QY	901	GTGGAGGAAGACTTTTGTGATGATGAACCAACTATCATTTAGAAGAGCTTTTACTCTTTGG	960
Db	901	GTGGAGGAAGACTTTTGTGATGATGAACCAACTATCATTTAGAAGAGCTTTTACTCTTTGG	960
QY	961	CAATTATATTGGGCTCCGAAATAGCTAAAGAACTGCGCAACTCTCTGTGATTGTAGA	1020
Db	961	CAATTATATTGGGCTCCGAAATAGCTAAAGAACTGCGCAACTCTCTGTGATTGTAGA	1020
QY	1021	CAGGTACTGGCAGACGCGCACCTATGCCATAGCCACTGAGGTGAGTGGGGTCTCCA	1080
Db	1021	CAGGTACTGGCAGACGCGCACCTATGCCATAGCCACTGAGGTGAGTGGGGTCTCCA	1080
QY	1081	GCACCTGCCCCAGCCCATCACCTGTGTACCAAGTGGCGCAGAGGACCTGCTCAAACTGA	1140
Db	1081	GCACCTGCCCCAGCCCATCACCTGTGTACCAAGTGGCGCAGAGGACCTGCTCAAACTGA	1140
QY	1141	CCTTATCTGCTGCTCACTGTGAGTCTGTGAGGAGAGGTTGCAGAGGCTCAGGGCGGGG	1200
Db	1141	CCTTATCTGCTGCTCACTGTGAGTCTGTGAGGAGAGGTTGCAGAGGCTCAGGGCGGGG	1200
QY	1201	CATGAGAGACACAGGGAAGACAGACTTGAGGCCAACAGTGTGTTTCTGTAAGGT	1260
Db	1201	CATGAGAGACACAGGGAAGACAGACTTGAGGCCAACAGTGTGTTTCTGTAAGGT	1260
QY	1261	AGAAATGTCTCTACCAGCGATGGAGATCTCGGCTGCCATGTGGTTGATGCCAGCCCCCTC	1320
Db	1261	AGAAATGTCTCTACCAGCGATGGAGATCTCGGCTGCCATGTGGTTGATGCCAGCCCCCTC	1320
QY	1321	CAGAGAAAAGTCTCTGCAGACAGTATTAAAGCCATAATCCAGAATAGTTTGTAGTAAACCGTA	1380
Db	1321	CAGAGAAAAGTCTCTGCAGACAGTATTAAAGCCATAATCCAGAATAGTTTGTAGTAAACCGTA	1380
QY	1381	GTTACTCTGGCGAGGTGCCACGTCTAACTAGATTAGATTGTTTGAACAATCTACATCC	1440
Db	1381	GTTACTCTGGCGAGGTGCCACGTCTAACTAGATTAGATTGTTTGAACAATCTACATCC	1440
QY	1441	ACCATTTGTATTGCAAGTGTCCCAAAATTTCTGTTCTACAAGCATGTGTGGCGCAAAA	1500
Db	1441	ACCATTTGTATTGCAAGTGTCCCAAAATTTCTGTTCTACAAGCATGTGTGGCGCAAAA	1500
QY	1501	CTGGAGACCAAGCACTTAAATTTTACTTCAGCCATCGTACCTCTCTGACTGATGGACC	1560
Db	1501	CTGGAGACCAAGCACTTAAATTTTACTTCAGCCATCGTACCTCTCTGACTGATGGACC	1560
QY	1561	CGTCACTCAAAAGTCCCTCTCATCATGTTCCAGTGAGAGGCCAGGATTTGCTTTCTTC	1620
Db	1561	CGTCACTCAAAAGTCCCTCTCATCATGTTCCAGTGAGAGGCCAGGATTTGCTTTCTTC	1620
QY	1621	TGGCATAGTAAACATTTTCTTGGAAATATGTTTCACTTAACTACATPACCAATATCTGGA	1680
Db	1621	TGGCATAGTAAACATTTTCTTGGAAATATGTTTCACTTAACTACATPACCAATATCTGGA	1680
QY	1681	AGACCTGTCTTACTCAGACAGCACAGGTTGTACAGAGCAGACAGATCTTCCAGAT	1740
Db	1681	AGACCTGTCTTACTCAGACAGCACAGGTTGTACAGAGCAGACAGATCTTCCAGAT	1740
QY	1741	CAGCAGGGAGACCCCGAGCCCTCTGCTTCTCTCTACACTGSCATGCTGATGAGATCTGAC	1800
Db	1741	CAGCAGGGAGACCCCGAGCCCTCTGCTTCTCTCTCTACACTGSCATGCTGATGAGATCTGAC	1800
QY	1801	ATGCCACACTTGGCTCTTCCACATCTGGTTGCACTCGTCACTGATGGGCTCGCTGCATCT	1860
Db	1801	ATGCCACACTTGGCTCTTCCACATCTGGTTGCACTCGTCACTGATGGGCTCGCTGCATCT	1860
QY	1861	CCCTCAGTCCCAAAATTTCTAGTAGCCAAAGTTCCTGTCAGAGGCTGTCTATGTGCTGGC	1920
Db	1861	CCCTCAGTCCCAAAATTTCTAGTAGCCAAAGTTCCTGTCAGAGGCTGTCTATGTGCTGGC	1920
QY	1921	TGCCCAAGGACACTCTCGACAGCCATTTTGGGTAAGGAACACTTACAAAGAGGCAT	1980
Db	1921	TGCCCAAGGACACTCTCGACAGCCATTTTGGGTAAGGAACACTTACAAAGAGGCAT	1980
QY	1981	TGATCTTGTGCTGAGGCTTCAGAGCCCTTTTGATAGGCTTCTGATCTCATTTCAAAAGAC	2040
Db	1981	TGATCTTGTGCTGAGGCTTCAGAGCCCTTTTGATAGGCTTCTGATCTCATTTCAAAAGAC	2040
QY	2041	ATTCAAGCCAGATGCTCCAACTGMAATPACCAACCTTCTGAAATTAATTTTGTCTT	2100
Db	2041	ATTCAAGCCAGATGCTCCAACTGMAATPACCAACCTTCTGAAATTAATTTTGTCTT	2100
QY	2101	ATTATATTTCTTTTCTTTTCTTAAAGAAATGGCTCTGAATAGAAATGCACATTTTCCA	2160
Db	2101	ATTATATTTCTTTTCTTTTCTTAAAGAAATGGCTCTGAATAGAAATGCACATTTTCCA	2160
QY	2161	CTGAACTCGATGCTATCATTTAGCCAAATCCAGTAATTTATTAATTAATCTATACAT	2220
Db	2161	CTGAACTCGATGCTATCATTTAGCCAAATCCAGTAATTTATTAATTAATCTATACAT	2220
QY	2221	AATATGTTTCTCAGCATAGGAGCTATGATTCAATTAATTAAGTGGAGTCAAAAGCTA	2280
Db	2221	AATATGTTTCTCAGCATAGGAGCTATGATTCAATTAATTAAGTGGAGTCAAAAGCTA	2280

QY 961 CAATTATATTTGGCTCCGAATAGCTAAAGATCTGCCAATCTCCTGTGATCTGAGA 1020
Db 961 CAATTATATTTGGCTCCGAATAGCTAAAGATCTGCCAATCTCCTGTGATCTGAGA 1020
QY 1021 CAGGTACTGGCAGACAGCGGCACCTATGCCATAGCACTGAGGTGAGTGGGGGTCTCCA 1080
Db 1021 CAGGTACTGGCAGACAGCGGCACCTATGCCATAGCACTGAGGTGAGTGGGGGTCTCCA 1080
QY 1081 GCACCTGCCCCAGCCCATCACCCCTGTGTACCACTGGCCAGAGGACCTGCTCAAACTGA 1140
Db 1081 GCACCTGCCCCAGCCCATCACCCCTGTGTACCACTGGCCAGAGGACCTGCTCAAACTGA 1140
QY 1141 CTTTATCTCTGCTCTCACTGTGAGTCTCTGAGGAGGTTGCGAGGCTCGAGGGCCGGG 1200
Db 1141 CTTTATCTCTGCTCTCACTGTGAGTCTCTGAGGAGGTTGCGAGGCTCGAGGGCCGGG 1200
QY 1201 CATGGAGAAGACAGGGAAGAGCAGAACTTGAAGCCACACAGTGTCTTCGTCAAAAGGT 1260
Db 1201 CATGGAGAAGACAGGGAAGAGCAGAACTTGAAGCCACACAGTGTCTTCGTCAAAAGGT 1260
QY 1261 AGAATGTCTTACGACGGATGAGAAATCTGTGCTGCCATGTGGTTGATGCCAGCCCTC 1320
Db 1261 AGAATGTCTTACGACGGATGAGAAATCTGTGCTGCCATGTGGTTGATGCCAGCCCTC 1320
QY 1321 CAGAGAAAAGGTCTCGACACAGTATTAAAGCCCTAATCCAGAAATGTTTGTAGTGAACCGTA 1380
Db 1321 CAGAGAAAAGGTCTCGACACAGTATTAAAGCCCTAATCCAGAAATGTTTGTAGTGAACCGTA 1380
QY 1381 GTTACTCTGGCCAGGTGCCACGCTTAACCTAGATTAGATGTTGTTGAAAACATCTACATCC 1440
Db 1381 GTTACTCTGGCCAGGTGCCACGCTTAACCTAGATTAGATGTTGTTGAAAACATCTACATCC 1440
QY 1441 ACCATTTGTTATGACGTGTTCCAAATTTCTGTTCTACAGCATGTTGTGTGCGAGAAA 1500
Db 1441 ACCATTTGTTATGACGTGTTCCAAATTTCTGTTCTACAGCATGTTGTGTGCGAGAAA 1500
QY 1501 CTGGAGACGAGCATCTTAATTTTACTTTCAGCCATCGTACCCCTCTCTGACTGATGGACC 1560
Db 1501 CTGGAGACGAGCATCTTAATTTTACTTTCAGCCATCGTACCCCTCTCTGACTGATGGACC 1560
QY 1561 CGTCACTACAAAAGTCCCTCTCATCATGTTCCAGTGAGAGGCCAGCGATTGCTTTCTTC 1620
Db 1561 CGTCACTACAAAAGTCCCTCTCATCATGTTCCAGTGAGAGGCCAGCGATTGCTTTCTTC 1620
QY 1621 TGGCATAGTAAACATTTCTTGGAAATATGTTTCACTTAATCACTACCAAAATATCTGGA 1680
Db 1621 TGGCATAGTAAACATTTCTTGGAAATATGTTTCACTTAATCACTACCAAAATATCTGGA 1680
QY 1681 AGACCTGTCTTACTCAGACAGCACAGGTGTACAGAAAGCAGACAGATCTCCAGAT 1740
Db 1681 AGACCTGTCTTACTCAGACAGCACAGGTGTACAGAAAGCAGACAGATCTCCAGAT 1740
QY 1741 CAGCAGGAGACCCCGAGCCTCTGCTTCTCTACACTGGCATGCTGATGAGATCGTGAC 1800
Db 1741 CAGCAGGAGACCCCGAGCCTCTGCTTCTCTACACTGGCATGCTGATGAGATCGTGAC 1800
QY 1801 ATGCCACATTTGGCTTCTTCCACATCTGGTTGACCTGTCATGATGGCTCGCTGATCT 1860
Db 1801 ATGCCACATTTGGCTTCTTCCACATCTGGTTGACCTGTCATGATGGCTCGCTGATCT 1860
QY 1861 CCTCAGTCCCAATTTCTAGTAGCCCAAGTCTCTGAGAGGCTGTATGTGCTCTGGC 1920
Db 1861 CCTCAGTCCCAATTTCTAGTAGCCCAAGTCTCTGAGAGGCTGTATGTGCTCTGGC 1920
QY 1921 TGCCCAAGGACACTCTCTGAGAGCCATTTTGGGTAAAGAAACATTAACAAAGAGCAT 1980
Db 1921 TGCCCAAGGACACTCTCTGAGAGCCATTTTGGGTAAAGAAACATTAACAAAGAGCAT 1980
QY 1981 TGATCTTGTCTGAGGCTCAGAGCCCTTTTGTAGGCTTCTGATGCTATTCATAAGAC 2040
Db 1981 TGATCTTGTCTGAGGCTCAGAGCCCTTTTGTAGGCTTCTGATGCTATTCATAAGAC 2040

QY 2041 ATTCAAGCCAGATGCTCCAACTGCAAAATATACCAACCTTCTCTGAATATATTTTGCCT 2100
Db 2041 ATTCAAGCCAGATGCTCCAACTGCAAAATATACCAACCTTCTCTGAATATATTTTGCCT 2100
QY 2101 ATTATATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2160
Db 2101 ATTATATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2160
QY 2161 TCTGAACGTGATGCATATCAITTAGCCAATCCAGTAATTTATATTTAACTATACAT 2220
Db 2161 TCTGAACGTGATGCATATCAITTAGCCAATCCAGTAATTTATATTTAACTATACAT 2220
QY 2221 AATATGTTTCTCCTCAGCATAGGAGCTATGATTCAATTAATAAAGTGGAGTCAAAACGCTA 2280
Db 2221 AATATGTTTCTCCTCAGCATAGGAGCTATGATTCAATTAATAAAGTGGAGTCAAAACGCTA 2280
QY 2281 AATGCAATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2340
Db 2281 AATGCAATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2340
QY 2341 CAAGTGGATCTTGGAGTGGATTTCTTGTTAAATTTCTTGCACTTGAATGTTCTATGAT 2400
Db 2341 CAAGTGGATCTTGGAGTGGATTTCTTGTTAAATTTCTTGCACTTGAATGTTCTATGAT 2400
QY 2401 TACATATGAAATCGCTTTGACATATCTTTAGACAGAAAAAGTAGCTGAGTGGGGGAA 2460
Db 2401 TACATATGAAATCGCTTTGACATATCTTTAGACAGAAAAAGTAGCTGAGTGGGGGAA 2460
QY 2461 ATTATAGACTGTGTGACTTTAGGAGTAGCTGCTCTTTATACATATCTCAAGCCCTG 2520
Db 2461 ATTATAGACTGTGTGACTTTAGGAGTAGCTGCTCTTTATACATATCTCAAGCCCTG 2520
QY 2521 AAGCCTTGATGCTCTCGACGCTCGCACTAAAGAGGGGCTTTTGCACC 2571
Db 2521 AAGCCTTGATGCTCTCGACGCTCGCACTAAAGAGGGGCTTTTGCACC 2571

RESULT 4

ID ADP84474 standard; DNA; 3060 BP.

XX ADP84474;

XX AC

XX 09-SEP-2004 (first entry)

XX Human breast-specific protein coding sequence #73.

XX human; breast-specific protein; breast cancer; gene; ds.

XX Homo sapiens.

XX WO2004053077-A2.

XX 24-JUN-2004.

XX 05-DEC-2003; 2003WO-US038815.

XX 05-DEC-2002; 2002US-0431123P.

XX (DIAD-) DIADEXUS INC.

XX Macina RA, Turner LR, Sun Y, Chen H, Rodriguez M;

XX WPI; 2004-468848/44.

XX P-PSDB; ADP84606.

XX New breast specific nucleic acid molecules and polypeptides useful for
PT diagnosing, preventing or treating breast cancer, for producing
transgenic animals or cells, or for research purposes.

XX Claim 1; SEQ ID NO 73; 521pp; English.

XX The invention comprises the amino acid and coding sequences of human

CC breast-specific proteins. The DNA and protein sequences of the invention
CC are useful for the diagnosis, treatment and prevention of breast cancer.
CC The present DNA sequence encodes a human breast-specific protein of the
CC invention.
XX
SQ Sequence 3060 BP; 731 A; 758 C; 757 G; 814 T; 0 U; 0 Other;

Query Match		92.6%;	Score 2380.6;	DB 12;	Length 3060;
Best Local Similarity		99.4%;	Pred. No. 0;		
Matches 2463;		Conservative	0;	Mismatches	9;
				Indels	7;
				Gaps	7;
QY	15	GGTGTCTGAGCGCCCATGGCTTCGCGCGCGCTCTGCGCGGCGCACTGTGGGGC	74		
DB	1	GGTGTCTGAGCGCCCATGGCTTCGCGCGCGCTCTGCGCGGCGCACTGTGGGGC	60		
QY	75	CGTGTCTGCGGCGCGGGGTCTGCGCTGGGCGCATGGCTCCGCGTCCGCTTCGTC	134		
DB	61	CGTGTCTGCGGCGCGGGGTCTGCGCTGGGCGCATGGCTCCGCGCGCGCTTCGTC	120		
QY	135	TGAGCTTCCGACTGCACTGCTGGCTCACTTGGCCCTAGGCGCGAGCGCCCGGGCAG	194		
DB	121	TGAGCTTCCGACTGCACTGCTGGCTCACTTGGCCCTAGGCGCGAGCGCCCGGGCAG	180		
QY	195	CAGACGCCCGGACCCCGCTGGCGGCGCTGTGGGGCGCCCGGAGCGAGCTACTCGC	254		
DB	181	CAGACGCCCGGACCCCGCTGGCGGCGCTGTGGGGCGCCCGGAGCGAGCTACTCGC	240		
QY	255	TGTGCTGCTGCTGACCCCGCGAGCGGCTGGGGCGCCGGGTCCGGGCGCGCGCTGC	314		
DB	241	TGTGCTGCTGCTGACCCCGCGAGCGGCTGGGGCGCCGGGTCCGGGCGCGCGCTGC	300		
QY	315	ACGAGGCTGCTGACACAGCTGCGCGCGCGCCCTTCCAGCGGTGCCAGCTGCTCAGC	374		
DB	301	ACGAGGCTGCTGACACAGCTGCGCGCGCGCCCTTCCAGCGGTGCCAGCTGCTCAGC	360		
QY	375	TGCTCTGCTACTGCTCGCGGCGCGGCGGCGGCGGCGGCGGCGGCTTCTGCTCGCG	434		
DB	361	TGCTCTGCTACTGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCTTCTGCTCGCG	420		
QY	435	ACCCCTTGATGACCTTGACACCGCGGAGCGGTGCTCGAGCTGCTGGCGCTGCCAG	494		
DB	421	ACCCCTTGATGACCTTGACACCGCGGAGCGGTGCTCGAGCTGCTGGCGCTGCCAG	480		
QY	495	AGGCACACCGCGGCTTGGGGAGTTGAGGCGGCGGCGGCGGCGGCGGCTGTCAGC	554		
DB	481	AGGCACACCGCGGCTTGGGGAGTTGAGGCGGCGGCGGCGGCGGCGGCTGTCAGC	540		
QY	555	GCTCTGGGAGGTGCAAGACGCGGCGGCTGCGAGTGGGCTGCGCACAGGTGCTGCCG	614		
DB	541	GCTCTGGGAGGTGCAAGACGCGGCGGCTGCGAGTGGGCTGCGCACAGGTGCTGCCG	600		
QY	615	TCCGAGCCCGCGTGCACCGGGTGGTGCCAGACTTGCACGTTCCGTTGCTTCCCGG	674		
DB	601	TCCGAGCCCGCGTGCACCGGGTGGTGCCAGACTTGCACGTTCCGTTGCTTCCCGG	660		
QY	675	ACGGGAACCGCGCGGCGGCTTGGAGAGTGTACCTCTTTATTCCTGAAGCCCGGG	734		
DB	661	ACGGGAACCGCGCGGCGGCTTGGAGAGTGTACCTCTTTATTCCTGAAGCCCGGG	720		
QY	735	CAGTGTCTGACCTGTGCAAGACGCGGCGGCTGCGAGTGGGCTGCGCACAGGTGCT	794		
DB	721	CAGTGTCTGACCTGTGCAAGACGCGGCGGCTGCGAGTGGGCTGCGCACAGGTGCT	780		
QY	795	TTGCCATCGAAGGACTGGATGCCACGGGTAAACACCGGTGACCCAGTCAGTGGCAGAT	854		
DB	781	TTGCCATCGAAGGACTGGATGCCACGGGTAAACACCGGTGACCCAGTCAGTGGCAGAT	840		
QY	855	CACCTAAGGCTGTCCTCTTAAGTACACCCCTCTTGATGAGTGGCCAGTGGAGAGATCT	914		
DB	841	CACCTAAGGCTGTCCTCTTAAGTACACCCCTCTTGATGAGTGGCCAGTGGAGAGATCT	900		
QY	915	TTGATGATGAACCAACTATCATTTAGAGAGCTTTTACTCTTTGGCAATTATTTGTGG	974		

DB	901	TTGATGATGAACCAACTATCATTTAGAGAGCTTTTACTCTTTGGCAATTATTTGTGG	960		
QY	975	CTCCGAAATAGCTAAAGAAATCTGCAAAATCTCTGTGATTTGTAGACAGGTACTGCA	1034		
DB	961	CTCCGAAATAGCTAAAGAAATCTGCAAAATCTCTGTGATTTGTAGACAGGTACTGCA	1020		
QY	1035	GCACGCCACCTATGCGCATAGCCACTGAGGTGAGTGGGGTCTCCAGCACCTGCCCCAG	1094		
DB	1021	GCACGCCACCTATGCGCATAGCCACTGAGGTGAGTGGGGTCTCCAGCACCTGCCCCAG	1080		
QY	1095	CCCATCACCTCTGTACACAGTGGCCAGAGGACTGCTCAAACTGACCTTATCTCTCTGC	1154		
DB	1081	CCCATCACCTCTGTACACAGTGGCCAGAGGACTGCTCAAACTGACCTTATCTCTCTGC	1140		
QY	1155	TCACCTGTGAGTCTGAGGAGAGTTGACAGGCTGCAGGCGCGGGGCATGGAAGACCA	1214		
DB	1141	TCACCTGTGAGTCTGAGGAGAGTTGACAGGCTGCAGGCGCGGGGCATGGAAGACCA	1200		
QY	1215	GGGAAGACGAACTTGAGGCGCAACAGTGTCTTTCGTCAAAAGGTAGAAATGCTCTACC	1274		
DB	1201	GGGAAGACGAACTTGAGGCGCAACAGTGTCTTTCGTCAAAAGGTAGAAATGCTCTACC	1260		
QY	1275	AGCGATGAGAAATCTCTGGCTGCCATGTGGTTGATGCCAGCCCTCCAGAAAAGGTCC	1334		
DB	1261	AGCGATGAGAAATCTCTGGCTGCCATGTGGTTGATGCCAGCCCTCCAGAAAAGGTCC	1320		
QY	1335	TCAGACAGTATTAAGCCTTAATCCAGATAGTTTGTAGTGAACCGTAGTTACTCTGGCCAG	1394		
DB	1321	TCAGACAGTATTAAGCCTTAATCCAGATAGTTTGTAGTGAACCGTAGTTACTCTGGCCAG	1380		
QY	1395	GTGCCACGCTTAACATAGATTAGATCTGTTTGAAACATCTATCCACCATTTGTATGC	1454		
DB	1381	GTGCCACGCTTAACATAGATTAGATCTGTTTGAACATCTATCCACCATTTGTATGC	1440		
QY	1455	AGTGTTCCTCAAAATTTCTGTTCTAAGCATGTTGTGGCAGAAAACCTGGAGACAGGCA	1514		
DB	1441	AGTGTTCCTCAAAATTTCTGTTCTAAGCATGTTGTGGCAGAAAACCTGGAGACAGGCA	1500		
QY	1515	TTCTAATTTTACTTCAGCCATCGTACCTCTTCTGATGATGAGCCGCTCATCAAAAGG	1574		
DB	1501	TTCTAATTTTACTTCAGCCATCGTACCTCTTCTGATGATGAGCCGCTCATCAAAAGG	1560		
QY	1575	TGCTCTCATCATGTTTCCAGTGAGGCGGCGAGTGTCTTCTCTGCGCATAGTAAACA	1634		
DB	1561	TGCTCTCATCATGTTTCCAGTGAGGCGGCGAGTGTCTTCTCTGCGCATAGTAAACA	1620		
QY	1635	TTTTCTTGAAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGTCTTACT	1694		
DB	1621	TTTTCTTGAAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGTCTTACT	1680		
QY	1695	CAGACAGCACAGGTGTACAGAAAGCAGCAGACAAGATCTTCAGATTCAGAGGAGACCC	1754		
DB	1681	CAGACAGCACAGGTGTACAGAAAGCAGCAGACAAGATCTTCAGATTCAGAGGAGACCC	1740		
QY	1755	CGGAGCCTCTGCTCTCTACAGCTGGCATGCTGATGAGATCGTGACATGCCCCACATTGGC	1814		
DB	1741	CGGAGCCTCTGCTCTCTACAGCTGGCATGCTGATGAGATCGTGACATGCCCCACATTGGC	1800		
QY	1815	TTCTTCCACATCTGGTTGCACTCGTATGATGGGCTCGCTGATCTCCCTCAGTCCCAAA	1874		
DB	1801	TTCTTCCACATCTGGTTGCACTCGTATGATGGGCTCGCTGATCTCCCTCAGTCCCAAA	1860		
QY	1875	TTCTAGTAGCCAAAGTGTCTCTCAGAGGCTGTCTATGTCTCTGGTCTCCCAAGGAGAC	1934		
DB	1861	TTCTAGTAGCCAAAGTGTCTCTCAGAGGCTGTCTATGTCTCTGGTCTCCCAAGGAGAC	1918		
QY	1935	TGCTGAGAGCCATTTTGGGTAAAGAACATCTAAGAAAGGCAATGATCTGTGTCTG	1994		
DB	1919	TGCTGAGAGCCATTTTGGGTAAAGAACATCTAAGAAAGGCAATGATCTGTGTCTG	1978		
QY	1995	AGGCTCAGAGCCCTTTTGTAGTGGCTTCTGATGCTCATTAATAAGACATTTCAAGCCAA	2054		
DB	1979	AGGCTCAGAGCCCTTTTGTAGTGGCTTCTGATGCTCATTAATAAGACATTTCAAGCCAA	2037		

Qy	2055	GCTCCAACTGCAAAATATACCAACTCTCTCGAAATTATATTTGCTTATTATATTCTTT	2114
Db	2038	GCTCCAACTGCAAAATATACCAACTCTCTCGAAATTATATTTGCTTATTATATTCTTT	2097
Qy	2115	TCCTTTTTTCTAAAGAAATGGCTCTGTAATAGAAATGCACATTTCCATCTCGAACTGGATGC	2174
Db	2098	TCCTTTTTTCTAAAGTA-TGGCTCTGAAATAGAAATGCACATTTTCCAT-TGGAATGGATGC	2155
Qy	2175	ATATCATTTAGCCAAATCCAGTAATTTATTTATTTATTAATCTATACATAAATGTTTCTCTCA	2234
Db	2156	ATTTTCATTTAGCCAAATCCAGTAATTTATTTATTTATTAATCTATACATAAATGTTTCTCTCA	2215
Qy	2235	GCATAGGAGCTATGATTCATTAAATTTAAAGTGGAGTCAAAACGCTAAATGCAATGTTGT	2294
Db	2216	GCATAGGAGCTATGATTCATTAAATTTAAAGTGGAGTCAAAACGCTAAATGCAATGTTGT	2275
Qy	2295	TGTGTATTTTCATTACACAAACTTAAATTTGCTCTTGTTAAATAGTCTCAAGTGAATCTGG	2354
Db	2276	TGTGTATTTTCATTACACAAACTTAAATTTGCTCTTGTTAAATAGTAC-AGTGGATCTGG	2334
Qy	2355	AGTCGGAAATTCCTGGTAAATTTATCTTCACCTTGAATGCTCATGATTACATATGAAATCG	2414
Db	2335	AGTCGGAAATTTCTTGGTAAATTTATCTTGCACTTGAATGCTCATGATTACATATGAAATCG	2394
Qy	2415	CTTTTGACATATCTTTTAGACAGAAAAAGTAGCTGAGTGAGGGGGAATTTAGAGCTTGT	2474
Db	2395	CTTTTGACATATCTTTTAGACAGAAAAAGTAGCTGAGTGAGGGGGAATTTAGAGC-TGT	2453
Qy	2475	GTGACTTTAGGAGTAGCT	2493
Db	2454	GTGACTTTAGGAGTAGCT	2472

RESULT 5	
AD140904	
ID	AD140904 standard; cDNA; 2907 BP.
XX	
XX	AC AD140904;
XX	
DT	22-APR-2004 (first entry)
XX	
DE	Human kinase and phosphatase KPP-17 encoding cDNA SEQ ID NO:70.
XX	
XX	human: kinase and phosphatase: KPP: enzyme; cardiovascular;
KW	antiarteriosclerotic; hypotensive; vasotropic; antiinflammatory;
KW	antianginal; anti-HIV; antiallergic; antiasthmatic; immunosuppressive;
KW	antithyroid; dermatological; antidiabetic; nephrotropic; antigout;
KW	gastrointestinal; neuroprotective; osteopathic; antiarthritic; uropathic;
KW	ophthalmological; antirheumatic; antiparkinsonian; nootropic;
KW	anticonvulsant; hepatotropic; CNS; antipsoriatic; haemostatic;
KW	cytostatic; antilipæmic; antiparasitic; antihelminthic; antibacterial;
KW	virucide; protozoacide; fungicide; gene therapy; kinase modulator;
KW	phosphatase modulator; cardiovascular disease; immune system disorder;
KW	neurological disorder; growth and development disorder;
KW	cell proliferative disorder; infection; gene: ss.

XX		
FH	Key	Location/Qualifiers
FT	CDS	17. .1366
FT		/tag= a
FT		/product= "kinase and phosphatase KPP-17"
XX		
PN	WO2004009778-A2.	
XX		
PD	29-JAN-2004.	
XX		
Pf	18-JUL-2003; 2003WO-US022650.	
XX		
PR	19-JUL-2002; 2002US-0397354P.	
PR	02-AUG-2002; 2002US-0400509P.	
PR	02-AUG-2002; 2002US-0400783P.	

15-AUG-2002; 2002US-0404027P.
(INCY-) INCYTE CORP.
Hafalia AJA, Emerling BM, Kable AE, Richardson TW, Becha SD;
Baughn MR, Tang YT, Lal PG, Lee SY, Griffin JA, Khare R;
Elliot VS, Jin P, Hawkins PR, Swarnakar A, Chawla NK, Tran UK;
Gururajan R, Ding L, Marquis JP, Thornton MB, Forsythe IJ, Lee EA;
Gietzen KJ, Ramkumar J;
WPI; 2004-132950/13.
P-PSDB; ADI40851.
New human kinases and phosphatases, useful for diagnosing, treating or preventing atherosclerosis, hypertension, AIDS, allergy, multiple sclerosis, osteoarthritis, Alzheimer's disease, Crohn's disease, cancer or hepatitis.
Claim 5; SEQ ID NO 70; 330pp; English.
The present sequence encodes a human kinase and phosphatase (KPP) protein. KPP sequences have cardiovascular, antiarteriosclerotic, hypotensive, vasotropic, antiinflammatory, antianginal, anti-HIV, antiallergic, antiasthmatic, immunosuppressive, antithyroid, dermatologic, antidiabetic, antiarthritic, uropathic, ophthalmological, neuroprotective, osteoporotic, antiarthritic, anticonvulsant, hepatotropic, antihemetic, antiparkinsonian, nootropic, antilipemic, antiparasitic, CNS, antipsoriatic, haemostatic, cytostatic, antilipemic and fungicide antihelminthic, antibacterial, virucide, protozoicide and fungicide activities, and can be used in gene therapy, and as kinase modulators and phosphatase modulators. KPP proteins, polynucleotides, agonists and antagonists can be used for diagnosing, treating or preventing disorders associated with aberrant expression of KPP, such as cardiovascular diseases (e.g. atherosclerosis, hypertension, vasculitis, angina pectoris or congestive heart failure), immune system disorders (e.g. AIDS, allergies, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout, irritable bowel syndrome, multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome or uveitis), neurological disorders (e.g. Alzheimer's disease, Parkinson's disease, Huntington's disease, dementia or epilepsy), disorders affecting growth and development (e.g. cirrhosis, hepatitis, mixed connective tissue disease, psoriasis or primary hypercholesterolaemia, hyperlipidaemia or cancer), or viral, bacterial, fungal, parasitic, protozoan or helminthic infections. The KPP and polynucleotides are also useful in assessing the effects of exogenous compounds on the expression of nucleic acids and kinases and phosphatases, or for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity and pharmacology. The polynucleotides encoding KPP are useful for creating transgenic animals to model human disease.
Sequence 2907 BP; 689 A; 739 C; 715 G; 764 T; 0 U; 0 Other;

	Query Match	89.0%;	Score 2288;	DB 12;	Length 2907;
	Best Local Similarity	98.2%;	Pred. No. 0;		
	Matches 2440;	Conservative 0;	Mismatches 30;	Indels 14;	Gaps 12
QY	17	GTGCTGAGGCGCCCATATGGCTTCGCGCGCGGTCTCTGCGCGGCCACTTCGCGGGCGG	76		
Db	2	GTGCTGAGGCGCCCATGGCTTCGCGCGCGGTCTCTGCGCGGCCACTTCGCGGGCGG	61		
QY	77	CTGCTCGGGCGGCGGGGCTCGGCTGCGGGGCATGCTCGCGTGCCTTCGTCCTG	136		
Db	62	CTGCTCGGGCGGCGGGGCTCGGCTGCGGGGCATGCTCGCGCGGCTTCGTCCTG	121		
QY	137	GAGCTTCCCGACTGACCCCTGGCTCACTTCGCCCTAGCGGCCGAGCCGCCGGACGCA	196		
Db	122	GAGCTTCCCGACTGACCCCTGGCTCACTTCGCCCTAGCGGCCGAGCCGCCGGACGCA	181		
QY	197	GACGCCCCGACCCCGCGCTCGCGCGCTGTGTGGGGCCCCCGGAGCGAGCTACTTCGCTG	256		

Db 182 GAGCGCCCGACCCGCCCTGGCGGGCGTGTGGGGCCCCGGAGCGGAGCTACTCGCGT 241
Qy 257 TCGGTGCCGTGACCCCGGACCCCGGTGCGGGGCCCGGGTCCGGGCGCGGGCTGCAC 316
Db 242 TCGGTGCCGTGACCCCGGACCCCGGTGCGGGGCCCGGGTCCGGGCGCGGGCTGCAC 301
Qy 317 CAGCGCGTCTGACCGAGTGTGCGCGCGGCCCTTCCAGGGTGCAGGTCTCAGGCTG 376
Db 302 CAGCGCGTCTGACCGAGTGTGCGCGCGGCCCTTCCAGGGTGCAGGTCTCAGGCTG 361
Qy 377 CTCTGTACTGCG 436
Db 362 CTCTGTACTGCG 421
Qy 437 CCGCTGGATGACCCCTGACCCCGGCAAGCGCTGCTCAGCTGCTGGCGGCCCTGCCAGGAG 496
Db 422 CCGCTGGATGACCCCTGACCCCGGCAAGCGCTGCTCAGCTGCTGGCGGCCCTGCCAGGAG 481
Qy 497 GCACACGCCCGCACTTGGCGGAGTTCGAGGCGGACCCCGGGCGCGCGCGCGCGCGCG 556
Db 482 GCACACGCCCGCACTTGGCGGAGTTCGAGGCGGACCCCGGGCGCGCGCGCGCGCGCG 541
Qy 557 CTCTGGAGGTGCAAGCG 616
Db 542 CTCTGGAGGTGCAAGCG 601
Qy 617 CCGGAGCCCCCGTGCACCCGCGGTGCAGACTTGCAGGTTCGCGTCTTCCGGAC 676
Db 602 CCGGAGCCCCCGTGCACCCGCGGTGCAGACTTGCAGGTTCGCGTCTTCCGGAC 661
Qy 677 CCGGAGCG 736
Db 662 CCGGAGCG 721
Qy 737 GTGCTTGACCTGTGACCAAGTGCACCAAGCAAGTGCAGGAGTTCAGGTTGTT 796
Db 722 GTGCTTGACCTGTGACCAAGTGCACCAAGCAAGTGCAGGAGTTCAGGTTGTT 781
Qy 797 GCATCGAAGGACTGGATGCCCGGTGAAACCAAGTGCAGGAGTTCAGGAGTTCAG 856
Db 782 GCATCGAAGGACTGGATGCCCGGTGAAACCAAGTGCAGGAGTTCAGGAGTTCAG 841
Qy 857 CTTAAGGCTGCTCTTAAGTGCACCAAGTGCAGGAGTTCAGGAGTTCAGGAGTTCAG 916
Db 842 CTTAAGGCTGCTCTTAAGTGCACCAAGTGCAGGAGTTCAGGAGTTCAGGAGTTCAG 901
Qy 917 GATGATGAACCAACTATCAATGAGAGGCTTTTACTCTTTGGGCAATTAATTTGGGC 976
Db 902 GATGATGAACCAACTATCAATGAGAGGCTTTTACTCTTTGGGCAATTAATTTGGGC 961
Qy 977 TCCGAAATAGCTAAAGAACTGCGCAATCTCTGTGANTGTAGACAGGTACTGGCACAG 1036
Db 962 TCCGAAATAGCTAAAGAACTGCGCAATCTCTGTGANTGTAGACAGGTACTGGCACAG 1021
Qy 1037 ACGGCCACCTATGCCATAGCCATGAGGTGAGTGGGGGTCTCCAGCAGCTGCCCGAGCC 1096
Db 1022 ACGGCCACCTATGCCATAGCCATGAGGTGAGTGGGGGTCTCCAGCAGCTGCCCGAGCC 1081
Qy 1097 CATCACCTGTGTACAGTGGCGCAGAGGACCTGCTCAAACTTACCTGCTGTGCTC 1156
Db 1082 CATCACCTGTGTACAGTGGCGCAGAGGACCTGCTCAAACTTACCTGCTGTGCTC 1141
Qy 1157 ACTGTGAGTCTGTAGAGAGGTTCAGAGGCTGCAGGCGCGGGGCAATGAGAGACCGAG 1216
Db 1142 ACTGTGAGTCTGTAGAGAGGTTCAGAGGCTGCAGGCGCGGGGCAATGAGAGACCGAG 1201
Qy 1217 GAAGAGCAGAACTTGAAGCCACAGTGTGTTTCGTCAAAAGTGAAGTTCCTACAG 1276
Db 1202 GAAGAGCAGAACTTGAAGCCACAGTGTGTTTCGTCAAAAGTGAAGTTCCTACAG 1261
Qy 1277 CGGATGGAAATCCTGGCTGCCATGTGGTTGATGCCAGCCCTCCAGAGAAAGGTCCTG 1336
Db 1262 CGGATGGAAATCCTGGCTGCCATGTGGTTGATGCCAGCCCTCCAGAGAAAGGTCCTG 1321

Qy 1337 CAGACAGTATTAAGCCTAATCCAGAAATAGTTTATGTGAACCGTAGTTACTCTGGCCAGGT 1396
Db 1322 CAGACGGTATTAAGCCTAATCCAGAAATAGTTTATGTGAACCGTAGTTACTCTGGCCAGGT 1381
Qy 1397 GCCAGCTCTAACTAGATTAGATTGTTTGAACATCTACATCCACCATTTGTTATGCAG 1456
Db 1382 GCCAGCTCTAACTAGATTAGATTGTTTGAACATCTACATCCACCATTTGTTATGCAG 1441
Qy 1457 TGTTCCTCAAAATTTCTGTTCTACAAGCATGTTGTGGCAGAAAACCTGGAGACCAAGCATC 1516
Db 1442 TGTTCCTCAAAATTTCTGTTCTACAAGCATGTTGTGGCAGAAAACCTGGAGACCAAGCATC 1501
Qy 1517 TTAATTTTATCTCAGCATCGTACCCCTCTCTGTGATGGACCCGTCATCAAAAGGTC 1576
Db 1502 TTAATTTTATCTCAGCATCGTACCCCTCTCTGTGATGGACCCGTCATCAAAAGGTC 1561
Qy 1577 CCTCTCATCATGTTCCAGTGAGAGCCAGCATGCTTTCTCTGGCATAGTAAACAT 1636
Db 1562 CCTCTCATCATGTTCCAGTGAGAGCCAGCATGCTTTCTCTGGCATAGTAAACAT 1621
Qy 1637 TTTCTTGGAAATATGTTTCACTTAATCACTACCAAAATATCTGGAAGACCTGTTACTCA 1696
Db 1622 TTTCTTGGAAATATGTTTCACTTAATCACTACCAAAATATCTGGAAGACCTGTTACTCA 1681
Qy 1697 GACAGCACAGGTGTACAGACGACAGACAGATCTTCCAGATCAGAGGAGACCCCG 1756
Db 1682 GACAGCACAGGTGTACAGACGACAGACAGATCTTCCAGATCAGAGGAGACCCCG 1741
Qy 1757 GAGCCTCTGCTTCTCTACACTGGCATGCTGATGAGATGCTGACATGCCACATTTGGCTT 1816
Db 1742 GAGCCTCTGCTTCTCTACACTGGCATGCTGATGAGATGCTGACATGCCACATTTGGCTT 1801
Qy 1817 CTTCCACATCTGTTTGCATCTGATGAGGCTGCTGATCTCCCTCAGTCCCAAAAT 1876
Db 1802 CTTCCACATCTGTTTGCATCTGATGAGGCTGCTGATCTCCCTCAGTCCCAAAAT 1861
Qy 1877 CTAGTAGCAAGTGTCTGAGAGGCTGTATGTGCTGCTGGCTGCCCAAGGACACTC 1936
Db 1862 CTAGTAGCAAGTGTCTGAGAGGCTGTATGTGCTGCTGGCTGCCCAAGGACACTC 1919
Qy 1937 CTGACAGAGCCTTTTGGTGAAGAACCTTACAAGAGGCAATTCAGTCTGCTGAG 1996
Db 1920 CTGACAGAGCCTTTTGGTGAAGAACCTTACAAGAGGCAATTCAGTCTGCTGAG 1979
Qy 1997 GCTCAGAGCCTTTTGTAGAGGCTCTGATGCTTCAATAAGACATTCAGGCAAGATGC 2056
Db 1980 GCTCAGAGCCTTTTGTAGAGGCTCTGATGCTTCAATAAGACATTCAGGCAAGATGC 2038
Qy 2057 TCCAACTGCAAAATATACCACTTCTGAAATTTATTTTGTATTTATTTCTTTTC 2116
Db 2039 TCCAACTGCAAAATATACCACTTCTGAAATTTATTTTGTATTTATTTCTTTTC 2098
Qy 2117 TTTTCTTCAAGAAATGCTCTGAAATGCAATTTCCATCTGAACCTGGATGCAT 2176
Db 2099 TTTTCTTCAAGAAATGCTCTGAAATGCAATTTCCATCTGAACCTGGATGCAT 2156
Qy 2177 ATCAATTTAGCCAAATCCAGTAAATTTATTTATTTATTTATTTATTTATTTCTCAGC 2236
Db 2157 TTTCAATTTAGCCAAATCCAGTAAATTTATTTATTTATTTATTTATTTATTTCTCAGC 2216
Qy 2237 ATAGAGCTATGATTTCAATT - AAAAGTGAGTCAAAACGCTAAATGCAATGTTTGT 2295
Db 2217 ATAGAGCTATGATTTCAATT - AAAAGTGAGTCAAAACGCTAAATGCAATGTTTGT - T 2275
Qy 2296 GTGTATTTTCAATACACAACTTATTTCTGTTGTTTAAATGATTTCAAGTGCATTTGGA 2355
Db 2276 GTGTATTTTCAATACACAACTTATTTCTGTTGTTTAAATGATTTCAAGTGCATTTGGA 2334
Qy 2356 GTGGGATTTCTT - GGTAAATTTATTT - - - GCACTTTGAATGCTCTCATGATTTACATATGAA 2411
Db 2335 GTGGGATTTCTTGGGTAAATTTATTTTCCCTTTGAAATGCTCTCTGATTTACATATGAA 2394

QY 503 CGCCGCACTTGGGCGAGTTCAGGCGCGACCGCGGGCGAGCTGTGGCAGCGCCTCTGG 562
DB 421 CGCCGCACTTGGGCGAGTTCAGGCGCGACCGCGGGCGAGCTGTGGCAGCGCCTCTGG 480
QY 563 GAGGTGCAAGACGGCAGCGGCTGCGAGGTGGCTGCGCACAGCTGCTGCCGCTCCGGAG 622
DB 481 GAGGTGCAAGACGGCAGCGGCTGCGAGGTGGCTGCGCACAGCTGCTGCCGCTCCGGAG 540
QY 623 CCCCCTGCAACCCGCTGTGCGAGACTTGGCCAGTTCCGTGGTCTTCCGACCGGGAA 682
DB 541 CCCCCTGCAACCCGCTGTGCGAGACTTGGCCAGTTCCGTGGTCTTCCGACCGGGAA 600
QY 683 GCGCCCGGGCGGTTTGGAGGAGTGTACTCTCTTTATTTCTTGAACCCGGGCGAGTCTT 742
DB 601 GCGCCCGGGCGGTTTGGAGGAGTGTACTCTCTTTATTTCTTGAACCCGGGCGAGTCTT 660
QY 743 GACCTGGTGCAGCCAGTGCACCAAGATCCAGAAAGGAAAGTTCCAGGTTTGTGCCATC 802
DB 661 GACCTGGTGCAGCCAGTGCACCAAGATCCAGAAAGGAAAGTTCCAGGTTTGTGCCATC 720
QY 803 GAAGGACTGGATGCCACG---GGTAAACCAACGGTGACCCAGTCAAGTGGCAGATTCACTT 859
DB 721 GAAGGACTGGATGCCACCGGTGGTAAACCAACGGTGACCCAGTCAAGTGGCAGATTCACTT 780
QY 860 AAGGCTGTCTCTTAAAGTCAACACCTCTTGCATTTGCCAGTGGAGGAGATCTTTGAT 919
DB 781 AAGGCTGTCTCTTAAAGTCAACACCTCTTGCATTTGCCAGTGGAGGAGATCTTTGAT 840
QY 920 GATGAACCAACTCATTTAGAAGAGCTTTTACTCTTTGGGCAATATATTTGTGGCTCC 979
DB 841 GATGAACCAACTCATTTAGAAGAGCTTTTACTCTTTGGGCAATATATTTGTGGCTCC 900
QY 980 GAAATAGCTAAAGATCTGCCAAATCTCTGTGATTTGTAGAGAGTACTGGCAGCAGCAG 1039
DB 901 GAAATAGCTAAAGATCTGCCAAATCTCTGTGATTTAGACA-----GGCACAGCAGC 954
QY 1040 GCCACTATGCATAGCCACTCAGGTGAGTGGGGTCTCCAGCACTGCCGCCAGCCCAT 1099
DB 955 GCCACTATGCATAGCCACTCAGGTGAGTGGGGTCTCCAGCACTGCCGCCAGCCCAT 1014
QY 1100 CACCTGTGTACAGTGGCCAGAGGACCTGCTCAACCTGACTTATCTGTGCTCACT 1159
DB 1015 CACCTGTGTACAGTGGCCAGAGGACCTGCTCAACCTGACTTATCTGTGCTCACT 1074
QY 1160 GTGAGTCTGAGAGAGGTTGAGAGGCTGCGAGGCGGGGAGTGAAGAACACAGGGAA 1219
DB 1075 GTGAGTCTGAGAGAGGTTGAGAGGCTGCGAGGCGGGGAGTGAAGAACACAGGGAA 1134
QY 1220 GAAGCAGAACTTGAGCCCAACAGTGTGTTGTCGTCAAAGGTAGAAATGCTTACCAGCG 1279
DB 1135 GAAGCAGAACTTGAGCCCAACAGTGTGTTGTCGTCAAAGGTAGAAATGCTTACCAGCG 1194
QY 1280 ATGGAGATCTTGGCTGCCATGTGTTGATGCCAGCCCTCCAGAGAAAGGTCTCTGCAG 1339
DB 1195 ATGGAGATCTTGGCTGCCATGTGTTGATGCCAGCCCTCCAGAGAAAGGTCTCTGCAG 1254
QY 1340 ACAGTATTAGCCTAATCCAGATAGTCTTGTAGAACCTAGTACTCTGCGCAGGTGCC 1399
DB 1255 ACGGTATTAGCCTAATCCAGATAGTCTTGTAGAACCTAGTACTCTGCGCAGGTGCC 1314
QY 1400 ACGTCTAAGCTAGATAGTGTGTTGAAACATCTACATCCACATTTGTTATGAGTGT 1459
DB 1315 ACGTCTAAGCTAGATAGTGTGTTGAAACATCTACATCCACATTTGTTATGAGTGT 1374
QY 1460 TCCCAATTTCTGTTTACAAGCATGTTGTGGCAGAAACCTGGAGACCGGCATCTTTA 1519
DB 1375 TCCCAATTTCTGTTTACAAGCATGTTGTGGCAGAAACCTGGAGACCGGCATCTTTA 1434
QY 1520 ATTTTACTTCAGCATCGTACCTCTTCTGATGATGGACCCGTCATCAACAAGGTCCCT 1579
DB 1435 ATTTTACTTCAGCATCGTACCTCTTCTGACTGATGGACCCGTCATCAACAAGGTCCCT 1494
QY 1580 CTCATCATGTTCCAGTGAGAGGCCAGGATGCTTTCTTCTGGCATAGTAACATTTTC 1639

DB 1495 CTCATCATGTTCCAGTGAGAGGCCAGGANTGCTTTCTTCCGCGATAGTAACATTTTC 1554
QY 1640 TTGGAACATATGTTTCACTTAATCACTCAAAATATCTGGAAGACCTGCTTACTCAGAC 1699
DB 1555 TTGGAACATATGTTTCACTTAATCACTCAAAATATCTGGAAGACCTGCTTACTCAGAC 1614
QY 1700 AGCACAGGTGTACAGAAGCAGACAGAGATCTTCCAGATCAGCAGGAGACCCCGGAG 1759
DB 1615 AGCACAGGTGTACAGAAGCAGACAGAGATCTTCCAGATCAGCAGGAGACCCCGGAG 1674
QY 1760 CCTCTGCTTCTCCTACACTTGCATGCTGATGAGATCTGACATGCCACATTTGGCTTCTT 1819
DB 1675 CCTCTGCTTCTCCTACACTTGCATGCTGATGAGATCTGACATGCCACATTTGGCTTCTT 1734
QY 1820 CCACATCTGGTTGCACTCGTCAATGATGGGCTCGCTGCACTCTCCCTCAGTCCCAAAATCTTA 1879
DB 1735 CCACATCTGGTTGCACTCGTCAATGATGGGCTCGCTGCACTCTCCCTCAGTCCCAAAATCTTA 1794
QY 1880 GTAGCAGAGTGTCTCTGACAGGCTGTCTATGTCTCTGGCTGCCAAGGAGACCTCTCTG 1939
DB 1795 G-AGCCAAAGTGTCTCTGACAGGCTGTCTATGTCTCTGGCTGGCCAA-GGACACTCTCTG 1852
QY 1940 CAGAGCCATTTTGGGTAAAGGAACACTTACAAGGAAGGCATTGATCTTGTGCTGAGGCT 1999
DB 1853 CAGAGCCATTTTGGGTAAAGGAACACTTACAAGGAAGGCATTGATCTTGTGCTGAGGCT 1912
QY 2000 CAGAGCCCTTTTGTAGGCTTCTGATGCTCAATTCATTAAGACATTTCAAGCCAAAGTCTCC 2059
DB 1913 CAGAGCCCTTTTGTAGGCTTCTGTA-GTCATATATAAGACATTTCAAGCCAAAGTCTCC 1971
QY 2060 AACTGCAATATACCAACCTTCTCTGAATATATTTTGTGCTTATTTATATTTCTTCTTCTT 2119
DB 1972 AACTGCAATATACCAACCTTCTCTGAATATATTTTGTGCTTATTTATATTTCTTCTTCTT 2031
QY 2120 TTTTCTAAAGAAATGGCTCTGAAATGACATGTCACATTTTCCATCTGAACCTGGATGCATATC 2179
DB 2032 TTTTCTAAAGTA-TGGCTCTGAATGAAATGCAATTTTCCAT-TGAACTGGATGCATATC 2089
QY 2180 ATTTAGCCAAATCCAGTAATTTATTTATATTAATCTATACATAATATGTTTCTCTCAGCATA 2239
DB 2090 ATTTAGCCAAATCCAGTAATTTATTTATATTAATCTATACATAATATGTTTCTCTCAGCATA 2149
QY 2240 GGAGCTATGATTCATTAATTAAGTGGAGTGCACAAACGCTAAATGCAATGTTTGTGTGT 2299
DB 2150 GGAGCTATGATTCATTAATTAAGTGGAGTGCACAAACGCTAAATGCAATGTTTGTGTGT 2209
QY 2300 ATTTTCTATTACACAACTTAATTTCTTGTGTTAAATAAGTTCGAAGTGGATCTTGGAGTGG 2359
DB 2210 ATTTTCTATTACACAACTTAATTTCTTGTGTTAAATAAGTAC-AGTGGATCTTGGAGTGG 2268
QY 2360 GATTTCTTGGTAAATTTATCTTGCACCTTGAATGCTCATGATTACATATGAAATCGCTTTG 2419
DB 2269 GATTTCTTGGTAAATTTATCTTGCACCTTGAATGCTCATGATTACATATGAAATCGCTTTG 2328
QY 2420 ACATATCTTTAGACAGAAAGTAGCTGAGTGGGGGGAATTTATAGAGCTTGTGTGAC 2479
DB 2329 ACATATCTTTAGACAGAAAGTAGCTGAGTGGGGGGAATTTATAGAGCTTGTGTGAC 2387
QY 2480 TTTAGGGAGTAGCT 2493
DB 2388 TTTAGGGAGTAGCT 2401

RESULT 7

ADC39135

ID ADC39135 standard; cDNA; 2856 BP.

XX ADC39135;

XX ADC39135;

DT 18-DEC-2003 (first entry)

XX

DE Novel human NOVX polypeptide coding sequence SEQ ID NO: 77.

XX ds; gene; antidiabetic; cytostatic; immunomodulator; anorectic;
KW antipalemic; nootropic; neuroprotective; immunostimulant;
KW antiparkinsonian; anti-HIV; antiasthmatic; antiinflammatory; hypotensive;
KW antiarteriosclerotic; hemostatic; osteopathic; gene therapy; NOVX;
KW diabetes; obesity; cancer; lymphoma; uterus cancer; prostate cancer;
KW dyslipidemia; anorexia; wasting disorder; Alzheimer's disease;
KW Parkinson's disorder; cachexia; cardiomyopathy; AIDS; asthma;
KW Crohn's disease; multiple sclerosis; hypertension; atherosclerosis;
KW hemophilia; graft-versus-host disease;
KW Albright hereditary osteodystrophy.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 28..1294
FT CDS /*tag= a
XX
XX WO2003010327-A2.
XX
XX 06-FEB-2003.
XX
XX 02-MAY-2002; 2002WO-US014199.
XX
XX 02-MAY-2001; 2001US-0288063P.
XX 03-MAY-2001; 2001US-0288395P.
XX 07-MAY-2001; 2001US-0289087P.
XX 09-MAY-2001; 2001US-0289817P.
XX 09-MAY-2001; 2001US-0289818P.
XX 11-MAY-2001; 2001US-0290194P.
XX 14-MAY-2001; 2001US-0290753P.
XX 15-MAY-2001; 2001US-0291181P.
XX 16-MAY-2001; 2001US-0291124P.
XX 18-MAY-2001; 2001US-0292001P.
XX 21-MAY-2001; 2001US-0292374P.
XX 22-MAY-2001; 2001US-0292587P.
XX 23-MAY-2001; 2001US-0293107P.
XX 25-MAY-2001; 2001US-0293747P.
XX 29-MAY-2001; 2001US-0294109P.
XX 29-MAY-2001; 2001US-0294110P.
XX 30-MAY-2001; 2001US-0294434P.
XX 31-MAY-2001; 2001US-0294827P.
XX 12-JUL-2001; 2001US-0304879P.
XX 31-JUL-2001; 2001US-0308901P.
XX 14-AUG-2001; 2001US-031270P.
XX 17-AUG-2001; 2001US-0313416P.
XX 10-SEP-2001; 2001US-0318463P.
XX 27-SEP-2001; 2001US-0325683P.
XX 18-OCT-2001; 2001US-0330292P.
XX 28-NOV-2001; 2001US-033873P.
XX 03-DEC-2001; 2001US-0336909P.
XX 03-DEC-2001; 2001US-0337552P.
XX 21-FEB-2002; 2002US-0359245P.
XX 01-MAY-2002; 2002US-00136826.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Miller CB, Kekuda R, Malyankar UM, Li L, Pena CE, Spytek KA;
PI Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zerhusen BD;
PI Patturajan M, Anderson DW, Mezes PS, Peyman JA, Macdougall JR;
PI Padigaru M, Raetelli L, Shenoy SG, Gerlach VL, Shimkets RA, Zhong M;
PI Edinger SR, Ellerman K;
XX
XX WPI; 2003-239445/23.
XX P-PSDB; ADC39136.
XX
XX New NOVX polypeptides and polynucleotides, useful in gene therapy,
XX particularly for treating or preventing a syndrome associated with a
XX human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
XX hypertension or hemophilia.
XX
XX Claim 20; SEQ ID NO 77; 748pp; English.

CC The invention relates to new isolated NOVX polypeptides, the genes
CC encoding them or sequences having at least 95% identity to the amino acid
CC or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,
CC particularly in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, which includes a pathology associated
CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for
CC treating, preventing or alleviating pathology associated with NOVX
CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
CC polypeptide are especially useful for treating or preventing e.g.
CC diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate
CC cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease,
CC Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
CC disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,
CC graft-versus-host disease or Albright hereditary osteodystrophy. The DNA
CC encoding the protein is useful in gene therapy for treating the above
CC conditions. These are also useful in developing powerful assay system for
CC functional analysis of various human disorders, as well as in diagnostic
CC applications. This sequence represents one of the NOVX genes of the
CC invention.
XX
XX Sequence 2856 BP; 700 A; 712 C; 692 G; 752 T; 0 U; 0 Other;
SQ
Query Match 88.5%; Score 2276; DB 10; Length 2856;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 10; Indels 16; Gaps 9;
QY 83 GGGCGGCGGGGTCTGCGCTGGGCGCATGGCTCCGCGTGGCGGCTTCTCTGGAGCTT 142
DB 1 GGGCGGCGGGGTCTGCGCTGGGCGCATGGCTCCGCGTGGCGGCTTCTCTGGAGCTT 60
QY 143 CCGGACTGCACCTGGCTCACTTCGCCCTAGGCGCGCGCCCGCGGACGACGCGCC 202
DB 61 CCGGACTGCACCTGGCTCACTTCGCCCTAGGCGCGCGCCCGCGGACGACGCGCC 120
QY 203 CCGGACCCCGCGCTGGCGGCTGTGGGCGCCCGGAGCGAGCTACTCGCTGTGGCTG 262
DB 121 CCGGACCCCGCGCTGGCGGCTGTGGGCGCCCGGAGCGAGCTACTCGCTGTGGCTG 180
QY 263 CCGGTGACCCCGGACGCGCGCTGGGCGCCCGGAGCGCGCGCTGACACGACGCGC 322
DB 181 CCGGTGACCCCGGACGCGCGCTGGGCGCCCGGAGCGCGCGCTGACACGACGCGC 240
QY 323 CTGCTGCACAGCTGCGCGCGCGCTTCCAGCGGTGCGAGCTGTCTAGGCTGTCTGTC 382
DB 241 CTGCTGCACAGCTGCGCGCGCGCTTCCAGCGGTGCGAGCTGTCTAGGCTGTCTGTC 300
QY 383 TACTGCCCGGCG 442
DB 301 TACTGCCCGGCG 360
QY 443 GATGACCTGCACACCGCGCAAGCGTGTGCGAGCTGTGGGCGCGCTGCCAGGAGGACCA 502
DB 361 GATGACCTGCACACCGCGCAAGCGTGTGCGAGCTGTGGGCGCGCTGTGAGGAGGACCA 420
QY 503 CGCCCGGCACTTGGGCGAGTTCGAGGCGCGACCCCGCGCGCGCGCGCGCGCGCGCGCG 562
DB 421 CGCCCGGCACTTGGGCGAGTTCGAGGCGCGACCCCGCGCGCGCGCGCGCGCGCGCG 480
QY 563 GAGTGCAAGACGCGGCGCGCGCTGAGTGGGTGGCGACAGGTCTGCGCGCGCGCGCGG 622
DB 481 GAGTGCAAGACGCGGCGCGCTGAGTGGGTGGCGACAGGTCTGCGCGCGCGCGCGG 540
QY 623 CCCCCTGTCACCCCGGTGGTGGCAGACTTCCCGAGTTCCTGTGTCTCCCGGACCGGAA 682
DB 541 CCCCCTGTCACCCCGGTGGTGGCAGACTTCCCGAGTTCCTGTGTCTCCCGGACCGGAA 600
QY 683 GCGCGCGCGCGCGTGTGGAGAGTGTACTCTCTTATTCCTGAAGCCCGGCGAGTCTT 742
DB 601 GCGCGCGCGCGCGTGTGGAGAGTGTACTCTCTTATTCCTGAAGCCCGGCGAGTCTT 660
QY 743 GACCTGTGACACGAGTCCCAACAGATCCAGAGAGGAAGTTCAGAGTGTGTGGCATC 802
DB 661 GACCTGTGACACGAGTCCCAACAGATCCAGAGAGGAAGTTCAGAGTGTGTGGCATC 720

QY 803 GAAGGCTGGATGCCACG---GGTAAACCAACGGTGACCCAGTCACTGGGAGATTCACTT 859
DB 721 GAAGGCTGGATGCCACGGGTGGTAAACCAACGGTGACCCAGTCACTGGGAGATTCACTT 780
QY 860 AAGGCTGTCTCTTAAAGTCAACACCTCTTGCAATGGCCAGTGGAGGAGATCTTTGAT 919
DB 781 AAGGCTGTCTCTTAAAGTCAACACCTCTTGCAATGGCCAGTGGAGGAGATCTTTGAT 840
QY 920 GATGAACCAACTATCATTTAGAAGAGCTTTTACTCTTTGGGCAATATATTGTGGCTCC 979
DB 841 GATGAACCAACTATCATTTAGAAGAGCTTTTACTCTTTGGGCAATATATTGTGGCTCC 900
QY 980 GAAATAGCTAAAGAAATCTGCCAAATCTCTGTGATTTGTAGACAGGTACTGGACACAGCG 1039
DB 901 GAAATAGCTAAAGAAATCTGCCAAATCTCTGTGATTTGTAGAC-----GGCACAGCAG 954
QY 1040 GCCACCTATGCCATAGCCACTGAGGTGAGTGGGGTCTCCAGCACCTGCCCCAGCCCAT 1099
DB 955 GCCACCTATGCCATAGCCACTGAGGTGAGTGGGGTCTCCAGCACCTGCCCCAGCCCAT 1014
QY 1100 CACCCCTGTACCAAGTGGCCAGAGGACCTGCTCAAACTGACCTTATCTGCTGCTCACT 1159
DB 1015 CACCCCTGTACCAAGTGGCCAGAGGACCTGCTCAAACTGACCTTATCTGCTGCTCACT 1074
QY 1160 CTGAGTCTCCTGAGGAGGTTGCAGAGGCTGCAGGGCCGGGGCATGGAGAAGACAGGGAA 1219
DB 1075 GTGAGTCTCCTGAGGAGGTTGCAGAGGCTGCAGGGCCGGGGCATGGAGAAGACAGGGAA 1134
QY 1220 GAAGCAGAACTTGGAGCCCAACAGTGTGTTTGTCTCAAAAGGTAGAAATGCTTACACAGCG 1279
DB 1135 GAAGCAGAACTTGGAGCCCAACAGTGTGTTTGTCTCAAAAGGTAGAAATGCTTACACAGCG 1194
QY 1280 ATGGAGAACTCTGGTGCATGTGTTGATGCCAGCCCTCCAGAGAAAGGTCTTGCAG 1339
DB 1195 ATGGAGAACTCTGGTGCATGTGTTGATGCCAGCCCTCCAGAGAAAGGTCTTGCAG 1254
QY 1340 ACAGTATTAAAGCCTAATCCAGATAGTCTTGTAGTGAACCGGTAGTACTTGGCCAGGTGCC 1399
DB 1255 ACGGTATTAAAGCCTAATCCAGATAGTCTTGTAGTGAACCGGTAGTACTTGGCCAGGTGCC 1314
QY 1400 ACGTCTAAGTATAGATGTGTTTGAACATCTACATCCACCTTCTGTATGCAAGTGT 1459
DB 1315 ACGTCTAAGTATAGATGTGTTTGAACATCTACATCCACCTTCTGTATGCAAGTGT 1374
QY 1460 TCCCAAAATCTCTTCTACAAGCATGTTGTGGCAGAAACCTGGAGACCAGGCATCTTA 1519
DB 1375 TCCCAAAATCTCTTCTACAAGCATGTTGTGGCAGAAACCTGGAGACCAGGCATCTTA 1434
QY 1520 ATTTTACTTTCAGGCATCGTACCTCTTCTGACTGATGGACCCGTATCAACAAGGTCCCT 1579
DB 1435 ATTTTACTTTCAGGCATCGTACCTCTTCTGACTGATGGACCCGTATCAACAAGGTCCCT 1494
QY 1580 CTCATCATGTTCCAGTGAGAGGCCAGCGATTGCTTCTTCTGGCATAGTAAACATTTTC 1639
DB 1495 CTCATCATGTTCCAGTGAGAGGCCAGCGATTGCTTCTTCTGGCATAGTAAACATTTTC 1554
QY 1640 TTGGAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGCTTACTCAGAC 1699
DB 1555 TTGGAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGCTTACTCAGAC 1614
QY 1700 AGCAGCAGTGTACAGAACAGCAGACAGATCTTCCAGATCAGCAGGAGACCCCGAG 1759
DB 1615 AGCAGCAGTGTACAGAACAGCAGACAGATCTTCCAGATCAGCAGGAGACCCCGAG 1674
QY 1760 CTTCTCTTCTCTCTACACTGGCATGCTGATGAGATCGTGACATGCCACATTTGGCTTCTT 1819
DB 1675 CTTCTCTTCTCTCTACACTGGCATGCTGATGAGATCGTGACATGCCACATTTGGCTTCTT 1734
QY 1820 CCAATCTCTGGTTCATCTGATGAGGTCTGCTGATCTCCCTCAGTCCCAAAATTTCTA 1879
DB 1735 CCAATCTCTGGTTCATCTGATGAGGTCTGCTGATCTCCCTCAGTCCCAAAATTTCTA 1794

QY 1880 GTAGCCAAAGTGTCTCTGACAGAGGCTGTCTATGTGTCTCTGGCTGCCCAAGGACACTCCTG 1939
DB 1795 G-AGCCAAAGTGTCTCTGACAGAGGCTGTCTATGTGTCTCTGGCTGCCCAA-GGACACTCCTG 1852
QY 1940 CAGAGCCATTTTGGTAAAGGAACACTTACAAGAGGCACTTGTCTGTCTGAGGCT 1999
DB 1853 CAGAGCCATTTTGGTAAAGGAACACTTACAAGAGGCACTTGTCTGTCTGAGGCT 1912
QY 2000 CAGAGCCCTTTGTATAGGCTTCTGTATGTCATTTCAATAAGACATTTCAAGCAAGATCCTC 2059
DB 1913 CAGAGCCCTTTGTATAGGCTTCTGTATGTCATTTCAATAAGACATTTCAAGCAAGATCCTC 1971
QY 2060 AACTGCAATATACCAACCTTCTCTGAATATATTTTGTATTTATATTTCTTTCTTT 2119
DB 1972 AACTGCAATATACCAACCTTCTCTGAATATATTTTGTATTTATATTTCTTTCTTT 2031
QY 2120 TTTTCTTAAAGAAATTTGCTCTGAATGCAATTTCCATCTGAATCTGAATCTGAATCTGAATCT 2179
DB 2032 TTTTCTTAAAGTA-TGGCTCTGAATGCAATTTCCAT-TGAACTTGAATGCAATTTTC 2089
QY 2180 ATTTAGCCAAATCCAGTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2239
DB 2090 ATTTAGCCAAATCCAGTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2149
QY 2240 GGAGCTATGATTTCAATTTAAAGTGGAGTCAAAACGCTAAATGCAATGTTTGTGT 2299
DB 2150 GGAGCTATGATTTCAATTTAAAGTGGAGTCAAAACGCTAAATGCAATGTTTGTGT 2209
QY 2300 ATTTTCAATACCAAACTTAATTTGTTCTTGTAAATTAAGTCAAGTGGATCTTGGAGTGG 2359
DB 2210 ATTTTCAATACCAAACTTAATTTGTTCTTGTAAATTAAGTCAAGTGGATCTTGGAGTGG 2268
QY 2360 GATTTCTTGGTAAATTTATCTTGCATGTAATCTCATGATTTACATATGAAATCGCTTTG 2419
DB 2269 GATTTCTTGGTAAATTTATCTTGCATGTAATCTCATGATTTACATATGAAATCGCTTTG 2328
QY 2420 ACATATCTTTAGACAGAAAAAGTAGCTGAGTGGGGGAAATTTATAGACTTGTGTGAC 2479
DB 2329 ACATATCTTTAGACAGAAAAAGTAGCTGAGTGGGGGAAATTTATAGACTTGTGTGAC 2387
QY 2480 TTTAGGGAGTAGCT 2493
DB 2388 TTTAGGGAGTAGCT 2401
RESULT 8
ADH42558
ID ADH42558 standard; DNA; 2856 BP.
XX ADH42558;
XX
DT 25-MAR-2004 (first entry)
DE Novel human nucleic acid NOV78a.
XX
XX ds; gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic;
KW anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective;
KW nootropic; antiparkinsonian; antiasthmatic; antiinfertility;
KW cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;
KW AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
XX tissue typing; preventive medicine; pharmacogenomic; vaccine.
OS Homo sapiens.
XX
XX WO2003102159-A2.
XX
PD 11-DEC-2003.
XX
PF 04-JUN-2003; 2003WO-US017573.
XX
PR 04-JUN-2002; 2002US-0385490P.
PR 04-JUN-2002; 2002US-0385615P.

PR 04-JUN-2002; 2002US-0385755P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386355P.
PR 06-JUN-2002; 2002US-0386357P.
PR 06-JUN-2002; 2002US-0386447P.
PR 06-JUN-2002; 2002US-0386459P.
PR 06-JUN-2002; 2002US-0386465P.
PR 06-JUN-2002; 2002US-0386864P.
PR 07-JUN-2002; 2002US-0386701P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0387078P.
PR 07-JUN-2002; 2002US-0387081P.
PR 07-JUN-2002; 2002US-0387083P.
PR 10-JUN-2002; 2002US-0387429P.
PR 10-JUN-2002; 2002US-0387540P.
PR 10-JUN-2002; 2002US-0387866P.
PR 11-JUN-2002; 2002US-0387606P.
PR 11-JUN-2002; 2002US-0387810P.
PR 11-JUN-2002; 2002US-0387859P.
PR 11-JUN-2002; 2002US-0387868P.
PR 11-JUN-2002; 2002US-0387896P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 12-JUN-2002; 2002US-0388432P.
PR 12-JUN-2002; 2002US-0388479P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389604P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390144P.
PR 19-JUN-2002; 2002US-0390209P.
PR 25-JUN-2002; 2002US-0391726P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402268P.
PR 12-AUG-2002; 2002US-0402822P.
PR 13-AUG-2002; 2002US-0403458P.
PR 15-AUG-2002; 2002US-0403617P.
PR 15-AUG-2002; 2002US-0403732P.
PR 26-AUG-2002; 2002US-0406182P.
PR 12-SEP-2002; 2002US-0410085P.
PR 13-SEP-2002; 2002US-0410505P.
PR 23-SEP-2002; 2002US-0412955P.
PR 30-SEP-2002; 2002US-0415195P.
PR 23-OCT-2002; 2002US-0420627P.
PR 23-OCT-2002; 2002US-0420718P.
PR 24-OCT-2002; 2002US-0420852P.
PR 31-OCT-2002; 2002US-0422750P.
PR 01-NOV-2002; 2002US-0423095P.
PR 05-NOV-2002; 2002US-0423748P.

(CURA-) CURAGEN CORP.

XX PA Alsobrook CE, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
XX PI Burgess JP, Casman SJ, Catterton E, Dhanabal M, Edinger SR;
PI Ellerman K, Etenberg S, Gangolli EA, Gerlach VL, Gorman L;
PI Grosse WM, Gunther E, Guo X, Gusev Y, Herrmann JL, Ji W, Kekuda R;
PI Khramtsov NV, Larochele WJ, Li L, Liang H, Low K, Macdougall JR;
PI MacLachlan T, Malyankar UN, McQueeney K, Mezick AJ, Miller CE;
PI Millet I, Padigar M, Patturajan M, Peyman JA, Qian X, Rastelli L;
PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;
PI Szytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
PI Wolenc AR, Zhong H, Zhong H;
XX WPI; 2004-053467/05.
DR P-PSDB; ADH42559.
XX

PT New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in
PT pharmacogenomics.

XX Claim 20; SEQ ID NO 1111; 1503pp; English.

XX The invention relates to 566 new isolated human polypeptides and their
XX encoding genes, sequences that are at least 95% identical to these or
XX sequences comprising one or more conservative substitutions in these. The
XX polypeptide, polynucleotide and antibodies against the polypeptides are
XX useful in diagnosing, treating or preventing NOVX-associated disorders,
XX e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
XX diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
XX Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
XX The nucleic acids are further used as hybridization probes, in chromosome
XX mapping, tissue typing, preventive medicine, and pharmacogenomics. The
XX polypeptides are also useful as vaccines. This sequence represents an
XX example of the nucleic acid sequence of the invention.

SQ Sequence 2856 BP; 700 A; 712 C; 692 G; 752 T; 0 U; 0 Other;

Query Match 88.5%; Score 2276; DB 12; Length 2856;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 10; Indels 16; Gaps 9;

QY 83 GGGCGCGCGGGTCTGCGCTGGGGCCATGGCTCCCGCTGCGCTGCTGCTGAGCTT 142
DB 1 GGGCGCGCGGGTCTGCGCTGGGGCCATGGCTCCCGCTGCGCTGCTGCTGAGCTT 60
QY 143 CCGACTGCACCTCGCTCACTTCGCTAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 202
DB 61 CCGACTGCACCTCGCTCACTTCGCTAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 203 CCGACCCCGCGCTGCGCGCGCTGTTGGGGCCCCCGAGGCGAGCTACTCGCTGGGTG 262
DB 121 CCGACCCCGCGCTGCGCGCGCTGTTGGGGCCCCCGAGGCGAGCTACTCGCTGGGTG 180
QY 263 CCGTGACCCCGCGCGCGCGCTGCGGGGCGCGGGTCCGGCGCGCGCGCGCGCGCGCG 322
DB 181 CCGTGACCCCGCGCGCGCGCTGCGGGGCGCGGGTCCGGCGCGCGCGCGCGCGCGCG 240
QY 323 CTGCTGCACAGCTGCGCGCGCGCTTCCAGCGTGCAGTGTCTCAGCTGTCTGCTGC 382
DB 241 CTGCTGCACAGCTGCGCGCGCGCTTCCAGCGTGCAGTGTCTCAGCTGTCTGCTGC 300
QY 383 TACTGCCCGGGCG 442
DB 301 TACTGCCCGGGCG 360
QY 443 GATGACCCCG 502
DB 361 GATGACCCCG 420
QY 503 CG 562
DB 421 CG 480
QY 563 GAGTGCAAGACCG 622
DB 481 GAGTGCAAGACCG 540
QY 623 CCG 682
DB 541 CCG 600
QY 683 CG 742
DB 601 CG 660
QY 743 GACCTGCTGACCG 802
DB 661 GACCTGCTGACCG 720

QY	803	GAAGGACTGGATGCCACG--GGTAAACCAACGGTGACCGAGTCATGTGGCAGATTCACATT	859
DB	721	GAAGGACTGGATGCCACGGGTGTGTAACACACCGTGACCGAGTCAGTCAGTCAGATTCACATT	780
QY	860	AAGGCTGTCTCTTAAAGTCACACACCTCTTGCAATTGGCGGAGTGGAGGAGATCTTTGAT	919
DB	781	AAGGCTGTCTCTTAAAGTCACACACCTCTTGCAATTGGCGGAGTGGAGGAGATCTTTGAT	840
QY	920	GATGAACCAATCATATTAGAAGAGCTTTTACTCTTTGGGCAATATATTGTGGCCTCC	979
DB	841	GATGAACCAATCATATTAGAAGAGCTTTTACTCTTTGGGCAATATATTGTGGCCTCC	900
QY	980	GAATAGCTAAAGAACTCGCAAACTCTCTGTGATTTGTAGACAGGTACTGGACACACG	1039
DB	901	GAATAGCTAAAGAACTCTCGCAAACTCTCTGTGATTTGTAGAC-----GGCACACACG	954
QY	1040	GCCACCTATGCCATAGCCATGAGGTGAGTGGGGGTCTCCAGCACCTGCCCGCCAGCCCAT	1099
DB	955	GCCACCTATGCCATAGCCATGAGGTGAGTGGGGGTCTCCAGCACCTGCCCGCCAGCCCAT	1014
QY	1100	CACCTCTGTACAGTGGCCAGAGGACCTGCTCAAACTTGACCTTATCTCTGCTGCTCACT	1159
DB	1015	CACCTCTGTACAGTGGCCAGAGGACCTGCTCAAACTTGACCTTATCTCTGCTGCTCACT	1074
QY	1160	GTGAGTCTCTGAGGAGAGGTTCAGAGGCTGCAGGGCCGGGGCATGGAAGACACAGGAA	1219
DB	1075	GTGAGTCTCTGAGGAGAGGTTCAGAGGCTGCAGGGCCGGGGCATGGAAGACACAGGAA	1134
QY	1220	GAAAGCAAACTTGAGCCCAACAGTGTGTTGCTCAAAAGGTAGAAATGCTTACACAGCG	1279
DB	1135	GAAAGCAAACTTGAGCCCAACAGTGTGTTGCTCAAAAGGTAGAAATGCTTACACAGCG	1194
QY	1280	ATGGAAATCTCTGGCTGCCATGTGGTTGATGCCAGCCCTCCAGAGAAAGTCTCTGCAG	1339
DB	1195	ATGGAAATCTCTGGCTGCCATGTGGTTGATGCCAGCCCTCCAGAGAAAGTCTCTGCAG	1254
QY	1340	ACAGTATTAAAGCTTAATCCAGAAATGTTTGTAGTAAACCGTAGTACTCTGGCCAGGTGC	1399
DB	1255	ACGGTATTAAAGCTTAATCCAGAAATGTTTGTAGTAAACCGTAGTACTCTGGCCAGGTGC	1314
QY	1400	ACGCTTAACCTAGATAGATGTTGTTTGAACATCTACATCCACCTTGTATGACGTGT	1459
DB	1315	ACGCTTAACCTAGATAGATGTTGTTTGAACATCTACATCCACCTTGTATGACGTGT	1374
QY	1460	TCCAAATTTCTGTTCTAAGCATGTTGTGGCAGAAACTGGAGACCGCATCTTA	1519
DB	1375	TCCAAATTTCTGTTCTAAGCATGTTGTGGCAGAAACTGGAGACCGCATCTTA	1434
QY	1520	ATTTTACTTTCAGGCATCGTACCTCTTCTGACTGATGGACCGGTATACAAAGGTCCCT	1579
DB	1435	ATTTTACTTTCAGGCATCGTACCTCTTCTGACTGATGGACCGGTATACAAAGGTCCCT	1494
QY	1580	CTCATCATGTTCCAGTGAGAGGCCAGCGATGCTTCTTCTGCGCATAGTAAACATTTTC	1639
DB	1495	CTCATCATGTTCCAGTGAGAGGCCAGCGATGCTTCTTCTGCGCATAGTAAACATTTTC	1554
QY	1640	TTGGAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGCTTACTCAGAC	1699
DB	1555	TTGGAACATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGCTTACTCAGAC	1614
QY	1700	AGCACAGGTGTACAGAACGAGCAGACAGATCTTCAGATAGCAGGGAGACCCCGAG	1759
DB	1615	AGCACAGGTGTACAGAACGAGCAGACAGATCTTCAGATAGCAGGGAGACCCCGAG	1674
QY	1760	CTCTCTCTTCTCTACACTGGCATGCTGATGATGCTGACATGCCACATTTGGCTTCTT	1819
DB	1675	CTCTCTCTTCTCTACACTGGCATGCTGATGATGCTGACATGCCACATTTGGCTTCTT	1734
QY	1820	CCACATCTGGTGTGACATCTGATGAGGTGGCTGCTCATCTCCCTCAGTCCCAAAATCTTA	1879
DB	1735	CCACATCTGGTGTGACATCTGATGAGGTGGCTGCTCATCTCCCTCAGTCCCAAAATCTTA	1794
QY	1880	GTAGCCCAAGTGTCTCTGCAGAGGCTGTCTATGTGTCTCTGGCTGCCCAAGGACACTCCTG	1939
DB	1795	G-AGCCCAAGTGTCTCTGCAGAGGCTGTCTATGTGTCTCTGGCTGCCCA-AGACACTCCTG	1852
QY	1940	CAGAGCCATTTTGGGTAAAGGAACCTTACAAGGAAGGCAATGATCTTGTGTCTGAGGCT	1999
DB	1853	CAGAGCCATTTTGGGTAAAGGAACCTTACAAGGAAGGCAATGATCTTGTGTCTGAGGCT	1912
QY	2000	CAGAGCCCTTTTGTATAGGCTTCTGTATGTCAATTATAAAGACATTTCAAGCAAGATCCTCC	2059
DB	1913	CAGAGCCCTTTTGTATAGGCTTCTGTATGTCAATTATAAAGACATTTCAAGCAAGATCCTCC	1971
QY	2060	AACTGCAAAATACCAACCTTCTCTGAATATATTTTGTCTTATTTATATTTCTTTCTTT	2119
DB	1972	AACTGCAAAATACCAACCTTCTCTGAATATATTTTGTCTTATTTATATTTCTTTCTTT	2031
QY	2120	TTTTCTTAAAGAAATTTGGCTCTGAATAGAAATGCACATTTTCCATCTGAATGCAATC	2179
DB	2032	TTTTCTTAAAGTA-TGGCTCTGAATAGAAATGCACATTTTCCAT-TGAATCTGGATGCAATTC	2089
QY	2180	ATTTAGCCAAATCCAGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	2239
DB	2090	ATTTAGCCAAATCCAGTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	2149
QY	2240	GGAGCTATGATTCATTAATTTAAAGTGGAGTCAAAACGCTAAATGCAATGTTTGTGTGT	2299
DB	2150	GGAGCTATGATTCATTAATTTAAAGTGGAGTCAAAACGCTAAATGCAATGTTTGTGTGT	2209
QY	2300	ATTTTCATTACACAACTTAATTTGCTTCTTGTAAATTAAGTTTCAAGTGGATCTTGGAGTGG	2359
DB	2210	ATTTTCATTACACAACTTAATTTGCTTCTTGTAAATTAAGTTTCAAGTGGATCTTGGAGTGG	2268
QY	2360	GAATTTCTGGTAAATTAATCTGCACCTGAAATGCTCATGATTTACATATGAAATCGCTTTG	2419
DB	2269	GAATTTCTGGTAAATTAATCTGCACCTGAAATGCTCATGATTTACATATGAAATCGCTTTG	2328
QY	2420	ACATATCTTTACACAGAAAAAGTAGCTGAGTGGGGGAAAAATTATAGAGCTTGTGTGAC	2479
DB	2329	ACATATCTTTACACAGAAAAAGTAGCTGAGTGGGGGAAAAATTATAGAGCTTGTGTGAC	2387
QY	2480	TTTAGGGAGTAGCT 2493	
DB	2388	TTTAGGGAGTAGCT 2401	
RESULT 9			
ADM10601			
ID	ADM10601	standard; cDNA; 2856 BP.	
XX	ADM10601;		
XX	20-MAY-2004	(first entry)	
XX	Human thymidine kinase cDNA.		
XX	Human; ss; gene; antisense gene therapy; WNT-7B;		
KW	N-acetylglucosaminyltransferase; voltage-gated K channel;		
KW	ion transport channel; Map3K8;		
KW	mitogen activated protein kinase kinase kinase 8; thymidine kinase;		
KW	H-ras; interleukin-1b; interleukin 8; cytosolic; antimicrobial;		
KW	antiinflammatory; cell proliferative disorder; infection; inflammation.		
OS	Homo sapiens.		
EH	Key	Location/Qualifiers	
FT	misc_feature	15..116	
FT		/*tag= a	
FT		/note= "Target region for antisense	
FT		oligonucleotides.Claimed in claim 16"	
FT		132..211	
FT		/*tag= b	
FT		/note= "Target region for antisense	
FT		oligonucleotides.Claimed in claim 16"	

[illegible]

CC recombinant protein from a sample. The nucleotides may be useful for
CC diagnosing, staging, treating, or monitoring the progression of treatment
CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion
CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale
CC genetic or gene expression analysis of several new nucleic acid
CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC diagnosing pre-pathologic disorders, and chronic or acute diseases
CC associated with abnormalities in the expression, amount or distribution
CC of the protein. The present sequence represents a cDNA of the invention
CC that is differentially expressed in activated vascular tissue. Note: The
CC sequence data for this patent did not form part of the specification, but
CC was obtained in electronic format directly from USPTO at
CC <http://seqdata.uspto.gov/sequence.html?docID=20020137081>
XX
SQ

Sequence 2645 BP; 671 A; 628 C; 622 G; 724 T; 0 U; 0 Other;

Query Match		82.7%;	Score 2125.2;	DB 8;	Length 2645;
Best Local Similarity		99.3%;	Pred. No. 0;		
Matches 2207;		Conservative	0;	Mismatches	8;
				Indels	7;
				Gaps	7;
QY	272	CCGAGCCGCGCTGCGGGGCCCGGGTCCGGGCGCGCGGTGCAACAGCGCCTGCTGAC	331		
DB	1	CCGAGCCGCGCTGCGGGGCCCGGGTCCGGGCGCGCGGTGCAACAGCGCCTGCTGAC	60		
QY	332	CAGCTGCGCGCGCGCGCTTCCAGCGGTGCACGTGCTCAGGCTGCTGCTACTGCCCG	391		
DB	61	CAGCTGCGCGCGCGCGCTTCCAGCGGTGCACGTGCTCAGGCTGCTGCTACTGCCCG	120		
QY	392	GGCGGCGCAGCGCGCGCGCACAGCAAGGCTTCTGTGCGCGCACCCCTGGATGACCT	451		
DB	121	GGCGGCGCAGCGCGCGCGCACAGCAAGGCTTCTGTGCGCGCACCCCTGGATGACCT	180		
QY	452	GACACCCGCGCAAGCGTCTCGAGCTGCTGGGCGCGCTGCCAGAGGCAACACCGCGCAC	511		
DB	181	GACACCCGCGCAAGCGTCTCGAGCTGCTGGGCGCGCTGCCAGAGGCAACACCGCGCAC	240		
QY	512	TTGGGCGAGTTCAGGCGCGACCGCGCGCGAGTGTGGCAGCGCTCTGGAGGTGCAA	571		
DB	241	TTGGGCGAGTTCAGGCGCGACCGCGCGCGAGTGTGGCAGCGCTCTGGAGGTGCAA	300		
QY	572	GACGCGAGCGGCTGCAAGTGGGCTGCGCACAGTCTGTGCGCGTCCCGGAGCCCGCGT	631		
DB	301	GACGCGAGCGGCTGCAAGTGGGCTGCGCACAGTCTGTGCGCGTCCCGGAGCCCGCGT	360		
QY	632	CACCGGTGTGCGCAGATTGCGCAGTTCGCGAGTCTTCCGCGACCGGGAAGCGCGCG	691		
DB	361	CACCGGTGTGCGCAGATTGCGCAGTTCGCGAGTCTTCCGCGACCGGGAAGCGCGCG	420		
QY	692	CCCGTTTGGAGGAGTGTACTCTTTTCTTGAAGCCCGCGCAGTGTGACCTGCTC	751		
DB	421	CCCGTTTGGAGGAGTGTACTCTTTTCTTGAAGCCCGCGCAGTGTGACCTGCTC	480		
QY	752	GACAGTGCACAAACAGATCCAGAAAGGAAAGTTCCAGTGTGTGCCATCGAAGGACTG	811		
DB	481	GACAGTGCACAAACAGATCCAGAAAGGAAAGTTCCAGTGTGTGCCATCGAAGGACTG	540		
QY	812	GATGCCACGGTAAAAACAGGTCACCGAGTCAGTGGCAGATTCACTTAAGCGTCTC	871		
DB	541	GATGCCACGGTAAAAACAGGTCACCGAGTCAGTGGCAGATTCACTTAAGCGTCTC	600		
QY	872	TTAAGTCAACCCCTCTGATTTGGCGCAGTGGAGGAGATCTTTGATGATGAACCAACT	931		
DB	601	TTAAGTCAACCCCTCTGATTTGGCGCAGTGGAGGAGATCTTTGATGATGAACCAACT	660		
QY	932	ATCATTAAGAGAGCTTTTACTCTTTGGGCAATATATTTGGCTCCGAAATAGCTAAA	991		
DB	661	ATCATTAAGAGAGCTTTTACTCTTTGGGCAATATATTTGGCTCCGAAATAGCTAAA	720		
QY	992	GAATCTGCAATCTCTCTGTGATTTGTAGACAGGTACTGGCAGCAGCGCCACCTATGCC	1051		
DB	721	GAATCTGCAATCTCTCTGTGATTTGTAGACAGGTACTGGCAGCAGCGCCACCTATGCC	780		

QY	1052	ATAGCCACTGAGGTGAGTGGGGTCTCCAGACACCTGCCCCAGCCCATCCCTGTGTATC	1111		
DB	781	ATAGCCACTGAGGTGAGTGGGGTCTCCAGACACCTGCCCCAGCCCATCCCTGTGTATC	840		
QY	1112	CAGTGGCCAGAGGACTGCTCAAACTGACCTTATCTGCTGCTGCTGCTGCTGCTGCTG	1171		
DB	841	CAGTGGCCAGAGGACTGCTCAAACTGACCTTATCTGCTGCTGCTGCTGCTGCTGCTG	900		
QY	1172	GAGAGTTCAGAGGCTGCGGGCGCGGATGGAGAACAGGAGGAGGAGGAGGAGGAGGAG	1231		
DB	901	GAGAGTTCAGAGGCTGCGGGCGCGGATGGAGAACAGGAGGAGGAGGAGGAGGAGGAG	960		
QY	1232	GAGGCCAACAGTGTGTTTCTCAAAAGGTAGAAATGCTCTACAGCGGATGGAGATCCT	1291		
DB	961	GAGGCCAACAGTGTGTTTCTCAAAAGGTAGAAATGCTCTACAGCGGATGGAGATCCT	1020		
QY	1292	GGCTGCATGTGGTTCATGCCAGCCCTCCAGAGAAAGGTCTCTGAGACAGTATTAAAGC	1351		
DB	1021	GGCTGCATGTGGTTCATGCCAGCCCTCCAGAGAAAGGTCTCTGAGACAGTATTAAAGC	1080		
QY	1352	CTAATCCAGAAATAGTGTGAAACCGTAGTTACTCTGGCCAGGTGCCAGCTAACTAG	1411		
DB	1081	CTAATCCAGAAATAGTGTGAAACCGTAGTTACTCTGGCCAGGTGCCAGCTAACTAG	1140		
QY	1412	ATTAGATGTTGTTGAACATCTACATCCACATTTGTTATGCAAGTGTCCCAAAATTTCT	1471		
DB	1141	ATTAGATGTTGTTGAACATCTACATCCACATTTGTTATGCAAGTGTCCCAAAATTTCT	1200		
QY	1472	GTTCTAAGCATGTTGTGGCAGAAACTGAGACACAGGATCTTAATTTTACTTTCTAG	1531		
DB	1201	GTTCTAAGCATGTTGTGGCAGAAACTGAGACACAGGATCTTAATTTTACTTTCTAG	1260		
QY	1532	CCATCTACCTCTCTGATGATGACCGCTCATCAAAAGGTCCCTCTCATCATGTTTC	1591		
DB	1261	CCATCTACCTCTCTGATGATGACCGCTCATCAAAAGGTCCCTCTCATCATGTTTC	1320		
QY	1592	CAGTGAAGGCCAGCATGCTTTCTTCTGGCATAGTAAACAATTTCTTTGGACATATG	1651		
DB	1321	CAGTGAAGGCCAGCATGCTTTCTTCTGGCATAGTAAACAATTTCTTTGGACATATG	1380		
QY	1652	TTTCACTTAATCTACTACCAATATCTGGAGACCTGCTTACTCAGACACACCATGTT	1711		
DB	1381	TTTCACTTAATCTACTACCAATATCTGGAGACCTGCTTACTCAGACACACCATGTT	1440		
QY	1712	ACAGAGCAGACACAGATCTTCCAGATCAGAGGGAGAGCCCGGAGCCTCTGCTTCTC	1771		
DB	1441	ACAGAGCAGACACAGATCTTCCAGATCAGAGGGAGAGCCCGGAGCCTCTGCTTCTC	1500		
QY	1772	CTACACTGGCATGCTGATGAGATGTCATGCCCAATTTGGCTTTCTTCCACATCTGGT	1831		
DB	1501	CTACACTGGCATGCTGATGAGATGTCATGCCCAATTTGGCTTTCTTCCACATCTGGT	1560		
QY	1832	GCATCTGTCATGATGGCTCGCTGCATCTCCCTCAGTCCCACAAATTTCTAGTAGCCAGT	1891		
DB	1561	GCATCTGTCATGATGGCTCGCTGCATCTCCCTCAGTCCCACAAATTTCTAGTAGCCAGT	1619		
QY	1892	TCCTGTCAGAGGCTGCTCTATGTCCTGGCTGCCCAAGGACACTCTCTCAGAGCCATTT	1951		
DB	1620	TCCTGTCAGAGGCTGCTCTATGTCCTGGCTGCCCAAGGACACTCTCTCAGAGCCATTT	1678		
QY	1952	TGGGTAAAGGAACACTTACAAAGAGGATTTGATCTTGTGCTGAGGCTCAGAGCCCTTT	2011		
DB	1679	TGGGTAAAGGAACACTTACAAAGAGGATTTGATCTTGTGCTGAGGCTCAGAGCCCTTT	1738		
QY	2012	GATAGGCTCTGATGTCATTAAGACATTTCAAGCCAGATCTCCAACTGCAAAATAT	2071		
DB	1739	GATAGGCTCTGATGTCATTAAGACATTTCAAGCCAGATCTCCAACTGCAAAATAT	1797		
QY	2072	ACCAACCTCTCTCAATATATTTTGTATTTATTTCTTTTCTTTTCTTTTCTTTTCTTAA	2131		
DB	1798	ACCAACCTCTCTCAATATATTTTGTATTTATTTTCTTTTCTTTTCTTTTCTTTTCTTAA	1857		
QY	2132	TTGGCTCTGAATAGATGCAATTTTCCATCTGAACTGAGTGCATATCATTTAGCCATC	2191		

Db 1858 -TGGCTCTGAATAGATGACATTTTCCAT-TGAACCTGGATGCAATTCATTAGCAATC 1915
QY 2192 CAGTAATTTATTTATTAATATATACATAATATGTTTCTCAGCATAGGAGCTATGATT 2251
Db 1916 CAGTAATTTATTTATTAATATATACATAATATGTTTCTCAGCATAGGAGCTATGATT 1975
QY 2252 CATTAAATTTAAAGTGGAGTCAAAAGCTAAATGCAATGTTTGTGTGATTTTCATTACA 2311
Db 1976 CATTAAATTTAAAGTGGAGTCAAAAGCTAAATGCAATGTTTGTGTGATTTTCATTACA 2035
QY 2312 CAAACTTAATTTGTTCTTGTAAATTAAGTTCAAGTGGATCTTGGAGTGGGATTTCTTGGA 2371
Db 2036 CAAACTTAATTTGTTCTTGTAAATTAAGTAC-AGTGGATCTTGGAGTGGGATTTCTTGGA 2094
QY 2372 AATTATCTTGCACTTGAATGTCATGATTACATATGAATCGCTTTGACATATCTTTAG 2431
Db 2095 AATTATCTTGCACTTGAATGTCATGATTACATATGAATCGCTTTGACATATCTTTAG 2154
QY 2432 ACAGAAAAAGTAGCTAGTGGAGGGGAAATATATAGAGCTTGTGACCTTTAGGAGTAG 2491
Db 2155 ACAGAAAAAGTAGCTAGTGGAGGGGAAATATATAGAGC-TGTGTGACCTTTAGGAGTAG 2213
QY 2492 CT 2493
Db 2214 GT 2215
RESULT 11
ADC39137
ID ADC39137 standard; cDNA; 2331 BP.
AC ADC39137;
XX
DT 18-DEC-2003 (first entry)
XX
DE Novel human NOVX polypeptide coding sequence SEQ ID NO: 79.
XX
KW ds; gene; antidiabetic; cytostatic; immunomodulator; anorectic;
KW antilipemic; nootropic; neuroprotective; immunostimulant;
KW antiparkinsonian; anti-HIV; antiasthmatic; antiinflammatory; hypotensive;
KW antiarteriosclerotic; hemostatic; osteopathic; gene therapy.; NOVX;
KW diabetes; obesity; cancer; lymphoma; uterus cancer; prostate cancer;
KW dyslipidemia; anorexia; wasting disorder; Alzheimer's disease;
KW Parkinson's disorder; cachexia; cardiomyopathy; AIDS; asthma;
KW Crohn's disease; multiple sclerosis; hypertension; atherosclerosis;
KW hemophilia; graft-versus-host disease;
KW Albright hereditary osteodystrophy.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 16..769
CDS /*tag= a
FN
PN WO2003010327-A2.
XX
PD 06-FEB-2003.
XX
PF 02-MAY-2002; 2002WO-US014199.
XX
PR 02-MAY-2001; 2001US-0288063P.
PR 03-MAY-2001; 2001US-0288395P.
PR 07-MAY-2001; 2001US-0289087P.
PR 09-MAY-2001; 2001US-0289817P.
PR 09-MAY-2001; 2001US-0289818P.
PR 11-MAY-2001; 2001US-0290194P.
PR 14-MAY-2001; 2001US-0290753P.
PR 15-MAY-2001; 2001US-0291181P.
PR 16-MAY-2001; 2001US-0291243P.
PR 18-MAY-2001; 2001US-0292001P.
PR 21-MAY-2001; 2001US-0292374P.
PR 22-MAY-2001; 2001US-0292587P.

PR 23-MAY-2001; 2001US-0293107P.
PR 25-MAY-2001; 2001US-0293747P.
PR 29-MAY-2001; 2001US-0294109P.
PR 29-MAY-2001; 2001US-0294110P.
PR 30-MAY-2001; 2001US-029434P.
PR 31-MAY-2001; 2001US-0294827P.
PR 12-JUL-2001; 2001US-0304879P.
PR 31-JUL-2001; 2001US-0308901P.
PR 14-AUG-2001; 2001US-0312270P.
PR 17-AUG-2001; 2001US-0313416P.
PR 10-SEP-2001; 2001US-0318463P.
PR 27-SEP-2001; 2001US-0325683P.
PR 18-OCT-2001; 2001US-0330292P.
PR 28-NOV-2001; 2001US-0333873P.
PR 03-DEC-2001; 2001US-0336909P.
PR 03-DEC-2001; 2001US-0337552P.
PR 21-FEB-2002; 2002US-0359245P.
PR 01-MAY-2002; 2002US-00136826.
XX
XX (CURA-) CURAGEN CORP.
PI Miller CE, Kekuda R, Malyankar UM, Li L, Pena CEA, Spytek KA;
PI Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zerhusen BD;
PI Patturajan M, Anderson DW, Mezes PS, Peyman JA, Macdougall JR;
PI Padigaru M, Rastelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;
PI Edinger SR, Ellerman K;
XX
XX WPI; 2003-239445/23.
XX P-PSDB; ADC39138.
XX
PT New NOVX polypeptides and polynucleotides, useful in gene therapy, a
PT particularly for treating or preventing a syndrome associated with a
PT human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
PT hypertension or hemophilia.
XX
PS Claim 20; SEQ ID NO 79; 748pp; English.
XX
CC The invention relates to new isolated NOVX polypeptides, the genes
CC encoding them or sequences having at least 95% identity to the amino acid
CC or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,
CC particularly in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, which includes a pathology associated
CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for
CC treating, preventing or alleviating pathology associated with NOVX
CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
CC polypeptide are especially useful for treating or preventing e.g.
CC diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate
CC cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease,
CC Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
CC disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,
CC graft-versus-host disease or Albrit hereditary osteodystrophy. The DNA
CC encoding the protein is useful in gene therapy for treating the above
CC conditions. These are also useful in developing powerful assay system for
CC functional analysis of various human disorders, as well as in diagnostic
CC applications. This sequence represents one of the NOVX genes of the
CC invention.
XX
SQ Sequence 2331 BP; 646 A; 502 C; 497 G; 686 T; 0 U; 0 Other;
Query Match 66.3%; Score 1704.8; DB 10; Length 2331;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 22; Indels 7; Gaps 7;
QY 670 CCCGACCGGGAAGCCGCCGCGTCTTGGAGAGGTGTACCTCTTTATTCCTGAAGC 729
Db 60 CTTGGCTCACTTCGCCCTAGCGCGTCTTGGAGAGGTGTACCTCTTTATTCCTGAAGC 119
QY 730 CCGGCACTGTCTGACCTGTGTCGACCACTGCCAAAACAGATCCAGAAAGAAATTCGA 789
Db 120 CCGGCACTGTCTGACCTGTGTCGACCACTGCCAAAACAGATCCAGAAAGAAATTCGA 179
QY 790 GGTGTTCCCATCGAAGGACTGGATCCACGGGTAAACACCGGTGACCCAGTCAGTCGC 849

Db 180 GGTGTTGGCATCGAAGGACTGGATGCCACGGTAAACCCAGGTGACCCAGTTCAGCGGC 239
Qy 850 AGATTCACTTAAGGTGTCTCTTAAGGTCAACCCCTCTTGCAATGGCCAGTGGAGGA 909
Db 240 AGATTCACTTAAGGTGTCTCTTAAGGTCAACCCCTCTTGCAATGGCCAGTGGAGGA 299
Qy 910 GATCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 969
Db 300 GATCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 359
Qy 970 TGTGSCCTCCGAAATAGCTAAAGAACTCTGCTGATGATGATGATGATGATGATGATGATG 1029
Db 360 TGTGSCCTCCGAAATAGCTAAAGAACTCTGCTGATGATGATGATGATGATGATGATGATG 419
Qy 1030 GCACAGCAGCGGCACCTATGSCCATAGCCACTGAGGTGAGTGGGGTCTCCAGCACTGCC 1089
Db 420 GCACAGCAGCGGCACCTATGSCCATAGCCACTGAGGTGAGTGGGGTCTCCAGCACTGCC 479
Qy 1090 CCCAGCCCATCACTCTGTACAGTGGCCAGAGGACCTGCTCAACCTGACCTTATCTCT 1149
Db 480 CCCAGCCCATCACTCTGTACAGTGGCCAGAGGACCTGCTCAACCTGACCTTATCTCT 539
Qy 1150 GTGTCTCACTGTGAGTCTGAGAGAGGTTGACAGGCTGACAGGCGCGGGGATGGAGAA 1209
Db 540 GTGTCTCACTGTGAGTCTGAGAGAGGTTGACAGGCTGACAGGCGCGGGGATGGAGAA 599
Qy 1210 GACAGGGGAAGAGCAGAACTTGAGGCCAACAGTGTGTTGTCACAAAGGTGAAATGTC 1269
Db 600 GACAGGGGAAGAGCAGAACTTGAGGCCAACAGTGTGTTGTCACAAAGGTGAAATGTC 659
Qy 1270 CTACAGCGGATGGAGAACTTGGCTGCTGATGATGATGATGATGATGATGATGATGATG 1329
Db 660 CTACAGCGGATGGAGAACTTGGCTGCTGATGATGATGATGATGATGATGATGATGATG 719
Qy 1330 GGTCTGCGACAGATTAAGCCCTAATCCAGAAATAGTTTGTAGTGAACCGTAGTACTCTG 1389
Db 720 GGTCTGCGACAGATTAAGCCCTAATCCAGAAATAGTTTGTAGTGAACCGTAGTACTCTG 779
Qy 1390 GCCAGGTGCCAGTCTTAACCTAGATGATGATGATGATGATGATGATGATGATGATGATG 1449
Db 780 GCCAGGTGCCAGTCTTAACCTAGATGATGATGATGATGATGATGATGATGATGATGATG 839
Qy 1450 TATGAGTGTCCCAAAATTCCTGTTCTACAGCATGTTGTGCGCAGAAACTGGAGACC 1509
Db 840 TATGAGTGTCCCAAAATTCCTGTTCTACAGCATGTTGTGCGCAGAAACTGGAGACC 899
Qy 1510 AGGCATCTTAATTTTACTTCAGGCATCGTACCCCTCTTCTGACATGATGGACCCGTCATC 1569
Db 900 AGGCATCTTAATTTTACTTCAGGCATCGTACCCCTCTTCTGACATGATGGACCCGTCATC 959
Qy 1570 AAGGTCCTCTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1629
Db 960 AAGGTCCTCTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1019
Qy 1630 ABAATTTTCTTGAAATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGTC 1689
Db 1020 ABAATTTTCTTGAAATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGTC 1079
Qy 1690 TTACTCAGACAGCAGCAGGTTGACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGGA 1749
Db 1080 TTACTCAGACAGCAGCAGGTTGACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGGA 1139
Qy 1750 GACCCCGAGGCTCTGCTTCTCTTACACTGGCATGCTGATGATGATGATGATGATGATGATG 1809
Db 1140 GACCCCGAGGCTCTGCTTCTCTTACACTGGCATGCTGATGATGATGATGATGATGATGATG 1199
Qy 1810 TTGCTCTTCTCCACATCTGTTGACATCTGATGAGGCTGCTGATGATGATGATGATGATGATG 1869
Db 1200 TTGCTCTTCTCCACATCTGTTGACATCTGATGAGGCTGCTGATGATGATGATGATGATG 1259
Qy 1870 CCAAAATCTAGTACCCAAAGTGTCTCTGACAGAGGCTGCTGATGATGATGATGATGATGATG 1929
Db 1260 CCAAAATCTAG-AGCCAAAGTGTCTCTGACAGAGGCTGCTGATGATGATGATGATGATGATG 1317

Qy 1930 GACACTCTGAGAGCCATTTTGGTAAAGAACACTTACAAAGAGGCAATGATCTTGT 1989
Db 1318 GACACTCTGAGAGCCATTTTGGTAAAGAACACTTACAAAGAGGCAATGATCTTGT 1377
Qy 1990 GTCTGAGGCTCAGAGCCCTTTTGTAGCTTCTGATGCTTCTGATGCTTCTGATGCTTCTG 2049
Db 1378 GTCTGAGGCTCAGAGCCCTTTTGTAGCTTCTGATGCTTCTGATGCTTCTGATGCTTCTG 1436
Qy 2050 AAGATGCTCCAACTGCAATATACCACTTCTCTGAACTTATTTGCTTATTTATATTT 2109
Db 1437 AAGATGCTCCAACTGCAATATACCACTTCTCTGAACTTATTTGCTTATTTATATTT 1496
Qy 2110 TCTTTCTTTTCTTAAAGAACTGGCTCTGAACTGAACTGCACTTTTCCATCTGAACCTG 2169
Db 1497 TCTTTCTTTTCTTAAAGAACTGGCTCTGAACTGAACTGCACTTTTCCAT-TGAACCTG 1554
Qy 2170 GATCATATCATTTAGCCAACTGCACTTATTTATTTATTTATTTATTTATTTATTTATTT 2229
Db 1555 GATCATATCATTTAGCCAACTGCACTTATTTATTTATTTATTTATTTATTTATTTATTT 1614
Qy 2230 CTTGAGCATGAGCTTATGATTTCAATTAATTAAGTGGAGTCAAAACGCTAAATGCAATG 2289
Db 1615 CTTGAGCATGAGCTTATGATTTCAATTAATTAAGTGGAGTCAAAACGCTAAATGCAATG 1674
Qy 2290 TTTGTTGCTTATTTTCAATTAACCAAACTTAAATTTGTTCTTTGTTTAAATTAAGTGGAT 2349
Db 1675 TTTGTTGCTTATTTTCAATTAACCAAACTTAAATTTGTTCTTTGTTTAAATTAAGTGGAT 1733
Qy 2350 CTTGAGTGGGATTTCTGTTAAATTTATTTGCACTTTGCACTTTGCACTTTGCACTTTGCA 2409
Db 1734 CTTGAGTGGGATTTCTGTTAAATTTATTTGCACTTTGCACTTTGCACTTTGCACTTTGCA 1793
Qy 2410 AATCGTTTGCATATCTTTAGACAGAAATAGTCTGAGTGGAGGAAATTTATAGAG 2469
Db 1794 AATCGTTTGCATATCTTTAGACAGAAATAGTCTGAGTGGAGGAAATTTATAGAG 1853
Qy 2470 CTTGTTGCTTATTTAGGAGTGTAGCT 2493
Db 1854 C-TGTTGCTTATTTAGGAGTGTAGCT 1876

RESULT 12

ADH42564

ID ADH42564 standard; DNA; 2331 BP.

XX ADH42564;

XX 25-MAR-2004 (first entry)

XX Novel human nucleic acid NOV78d.

XX db, gene; cardiovascular; antiarteriosclerotic; hypotensive; cytostatic;
XX anorectic; antidiabetic; immunosuppressive; anti-HIV; neuroprotective;
XX KW cardiotropic; antiparkinsonian; antiasthmatic; antinfertility;
XX KW cardiomyopathy; atherosclerosis; hypertension; cancer; obesity; diabetes;
XX AIDS; multiple sclerosis; graft-versus-host disease; Alzheimer's disease;
XX KW Parkinson's disease; asthma; fertility disorder; chromosome mapping;
XX tissue typing; preventive medicine; pharmacogenomic; vaccine.

XX Homo sapiens.

XX WO2003102159-A2.

XX 11-DEC-2003.

XX 04-JUN-2003; 2003WO-US017573.

XX 04-JUN-2002; 2002US-0385490P.

XX 04-JUN-2002; 2002US-0385615P.

XX 04-JUN-2002; 2002US-0385755P.

XX 05-JUN-2002; 2002US-0386041P.

XX 06-JUN-2002; 2002US-0386355P.

PR 06-JUN-2002; 2002US-0386357P.
 PR 06-JUN-2002; 2002US-0386447P.
 PR 06-JUN-2002; 2002US-0386459P.
 PR 06-JUN-2002; 2002US-0386465P.
 PR 06-JUN-2002; 2002US-0386854P.
 PR 07-JUN-2002; 2002US-0386701P.
 PR 07-JUN-2002; 2002US-0386796P.
 PR 07-JUN-2002; 2002US-0386931P.
 PR 07-JUN-2002; 2002US-0387078P.
 PR 07-JUN-2002; 2002US-0387081P.
 PR 07-JUN-2002; 2002US-0387083P.
 PR 10-JUN-2002; 2002US-0387429P.
 PR 10-JUN-2002; 2002US-0387540P.
 PR 10-JUN-2002; 2002US-0387866P.
 PR 11-JUN-2002; 2002US-0387966P.
 PR 11-JUN-2002; 2002US-0387606P.
 PR 11-JUN-2002; 2002US-0387610P.
 PR 11-JUN-2002; 2002US-0387659P.
 PR 11-JUN-2002; 2002US-0387668P.
 PR 11-JUN-2002; 2002US-0387696P.
 PR 11-JUN-2002; 2002US-0387859P.
 PR 12-JUN-2002; 2002US-0387934P.
 PR 12-JUN-2002; 2002US-0387960P.
 PR 12-JUN-2002; 2002US-0388022P.
 PR 12-JUN-2002; 2002US-0388096P.
 PR 12-JUN-2002; 2002US-0388432P.
 PR 12-JUN-2002; 2002US-0388479P.
 PR 13-JUN-2002; 2002US-0389123P.
 PR 14-JUN-2002; 2002US-0389120P.
 PR 14-JUN-2002; 2002US-0389120P.
 PR 14-JUN-2002; 2002US-0389146P.
 PR 17-JUN-2002; 2002US-0389742P.
 PR 18-JUN-2002; 2002US-0389604P.
 PR 18-JUN-2002; 2002US-0389884P.
 PR 19-JUN-2002; 2002US-0390006P.
 PR 19-JUN-2002; 2002US-0390144P.
 PR 19-JUN-2002; 2002US-0390209P.
 PR 25-JUN-2002; 2002US-0391726P.
 PR 06-AUG-2002; 2002US-0401628P.
 PR 09-AUG-2002; 2002US-0402268P.
 PR 12-AUG-2002; 2002US-0402822P.
 PR 13-AUG-2002; 2002US-0403458P.
 PR 15-AUG-2002; 2002US-0403617P.
 PR 15-AUG-2002; 2002US-0403732P.
 PR 26-AUG-2002; 2002US-0406182P.
 PR 12-SEP-2002; 2002US-0410085P.
 PR 13-SEP-2002; 2002US-0410505P.
 PR 23-SEP-2002; 2002US-0412955P.
 PR 30-SEP-2002; 2002US-0415195P.
 PR 23-OCT-2002; 2002US-0420627P.
 PR 23-OCT-2002; 2002US-0420718P.
 PR 24-OCT-2002; 2002US-0420852P.
 PR 31-OCT-2002; 2002US-0422750P.
 PR 01-NOV-2002; 2002US-0423095P.
 PR 05-NOV-2002; 2002US-0423748P.
 (CURA-) CURAGEN CORP.
 PI Alsbrook JP, Anderson DW, Baumgartner JC, Berghs C, Boldog FL;
 PI Burgess CE, Casman SU, Catterton E, Dhanabal M, Edinger SR;
 PI Ellerman K, Ettenberg S, Gangolli EA, Gerlach VL, Gorman L;
 PI Grosse WM, Gunther E, Guo X, Gusev VY, Herrmann JL, Ji W, Kekuda R;
 PI Khrantsov NV, Larochelel WJ, Li L, Liang H, Low K, Macdougall JR;
 PI MacLachlan T, Malyankar UM, McQueeney K, Mezick AJ, Miller CB;
 PI Miller I, Padigar M, Patturajan M, Peyman JA, Qian X, Rastelli L;
 PI Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA, Smithson G;
 PI Spytek KA, Stone DJ, Sukumaran S, Szekeres ES, Vernet CAM, Voss EZ;
 PI Wolenc AR, Zhong M, Zhong H;
 WPI: 2004-053467/05.
 DR P-PSDB; ADH42565.
 DR
 XX
 PR New NOVX polypeptides and nucleic acid molecules useful for preventing or
 PR treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
 PT atherosclerosis or diabetes, in chromosome mapping, tissue typing or in

PT pharmacogenomics.
 XX
 PS Claim 20; SEQ ID NO 1117; 1503pp; English.
 XX
 CC The invention relates to 566 new isolated human polypeptides and their
 CC encoding genes, sequences that are at least 95% identical to these or
 CC sequences comprising one or more conservative substitutions in these. The
 CC polypeptide, polynucleotide and antibodies against the polypeptides are
 CC useful in diagnosing, treating or preventing NOVX-associated disorders,
 CC e.g. cardiomyopathy, atherosclerosis, hypertension, cancer, obesity,
 CC diabetes, AIDS, multiple sclerosis, graft-versus-host disease,
 CC Alzheimer's disease, Parkinson's disease, asthma, or fertility disorders.
 CC The nucleic acids are further used as hybridization probes, in chromosome
 CC mapping, tissue typing, preventive medicine, and pharmacogenomics. The
 CC polypeptides are also useful as vaccines. This sequence represents an
 CC example of the nucleic acid sequence of the invention.
 XX
 SQ Sequence 2331 BP; 646 A; 502 C; 497 G; 686 T; 0 U; 0 Other;
 Query Match 66.3%; Score 1704.8; DB 12; Length 2331;
 Best Local Similarity 98.4%; Pred. No. 0;
 Matches 1795; Conservative 0; Mismatches 22; Indels 7; Gaps 7;
 QY 670 CCGGACCGGGAAGCCCGCGCGCGCTTTTGGAGGAGTGCTCTCTTTATTCCTGAAGC 729
 Db 60 CTTGGCTCACTTCGCCCTAGCGCGCGCTTTGGAGGAGTGCTCTCTTTATTCCTGAAGC 119
 QY 730 CCGGCGAGTGCCTTGACCTGCTGACCAAGTGCCTCAAAACAGATCCAGAAAGGAAGTTCCA 789
 Db 120 CCGGCGAGTGCCTTGACCTGCTGACCAAGTGCCTCAAAACAGATCCAGAAAGGAAGTTCCA 179
 QY 790 GGTGTTGCCATCGAAGGACTGATGCCACGGGTAAACACCGGTGACCCAGTCACTGGC 849
 Db 180 GGTGTTGCCATCGAAGGACTGATGCCACGGGTAAACACCGGTGACCCAGTCACTGGC 239
 QY 850 AGATTCACCTTAAGGCTGCTCTTAAGTCAACACCTCTTCATTCGGCAGTGGAGAA 909
 Db 240 AGATTCACCTTAAGGCTGCTCTTAAGTCAACACCTCTTCATTCGGCAGTGGAGAA 299
 QY 910 GATCTTTGATGATGAACCAACTATCATTAAGAGAGCTTTTACTCTTTGGCAATTATAT 969
 Db 300 GATCTTTGATGATGAACCAACTATCATTAAGAGAGCTTTTACTCTTTGGCAATTATAT 359
 QY 970 TGTGGCTCCGAAATAGCTTAAAGATCTGCCAAATCTCTGTGATTTAGACAGGTACTG 1029
 Db 360 TGTGGCTCCGAAATAGCTTAAAGATCTGCCAAATCTCTGTGATTTAGACAGGTACTG 419
 QY 1030 GCACGACGCGCCACTATGCCATAGCCACTGAGGTGAGTGGGGTCTCCAGCACCTGCC 1089
 Db 420 GCACGACGCGCCACTATGCCATAGCCACTGAGGTGAGTGGGGTCTCCAGCACCTGCC 479
 QY 1090 CCCAGCCCATCACCTGTGTACAGTGGCCAGAGGACCTGCTCAAAACCTGACCTTATCCT 1149
 Db 480 CCCAGCCCATCACCTGTGTGTACAGTGGCCAGAGGACCTGCTCAAAACCTGACCTTATCCT 539
 QY 1150 GGTGCTCATGTGAGTCTTGAGAGAGGTTCAGAGGCTGCAGGGCGCGGGCATGAGAA 1209
 Db 540 GGTGCTCATGTGAGTCTTGAGAGAGGTTCAGAGGCTGCAGGGCGCGGGCATGAGAA 599
 QY 1210 GACCCGGGGAAGCAGACACTTGGGCCCAACAGTGTGTTTCCTCAAAAGGTAGAATGTC 1269
 Db 600 GACCCGGGGAAGCAGACACTTGGGCCCAACAGTGTGTTTCCTCAAAAGGTAGAATGTC 659
 QY 1270 CTACGAGCGGATGGAGAATCTCTGGTGCCTATGTTGATGTCAGCCCTCCAGAGAAAA 1329
 Db 660 CTACGAGCGGATGGAGAATCTCTGGTGCCTATGTTGATGTCAGCCCTCCAGAGAAAA 719
 QY 1330 GGTCTTCGACAGTATTAAAGCCTAATCCAGAAATAGTTTTAGTGAACCGTAGTACTCTG 1389
 Db 720 GGTCTTCGACAGTATTAAAGCCTAATCCAGAAATAGTTTTAGTGAACCGTAGTACTCTG 779
 QY 1390 GCCAGGTGCCAGTCTAATAGATTAGATTGTTGTTGAAACATCTACATCCACCATTTGT 1449

Db	780	GCACGGTGGCCACGCTCTAACTTAGATAGATGTTGTTGAAACATCTACATCCACCAATTGTT	839
Qy	1450	TATGCAGTGTTCCTCAAAATTCCTGTTCTCAACGATGTTGTGTGGCAGAAAACTGGAGACC	1509
Db	840	TATGCAGTGTTCCTCAAAATTCCTGTTCTCAACGATGTTGTGTGGCAGAAAACTGGAGACC	899
Qy	1510	AGGCATCTTAATTTTACTTTCAGCCATCGTACCCCTCTTCTGACTGATGAGACCGTCAATCAC	1569
Db	900	AGGCATCTTAAGTTTACTTCAGCCATCGTACCCCTCTTCTGACTGATGAGACCGTCAATCAC	959
Qy	1570	AAAGGTCCCTCTCATCATGTTCCAGTGAGAGGCCAGCGATTGCTTCTTCCTGGCATAGT	1629
Db	960	AAAGGTCCCTCTCATCATGTTCCAGTGAGAGGCCAGCGATTGCTTCTTCCTGGCATAGT	1019
Qy	1630	AAACATTTTCTTGGAAATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGTC	1689
Db	1020	AAACATTTTCTTGGAAATATGTTTCACTTAATCACTACCAATATCTGGAAGACCTGTC	1079
Qy	1690	TTACTTCAGACAGCACCAAGGTGTACAGAAAGCAGCAGACAAGATCTTCAGATCAGACGGGA	1749
Db	1080	TTACTTCAGACAGCACCAAGGTGTACAGAAAGCAGCAGACAAGATCTTCAGATCAGACGGGA	1139
Qy	1750	GACCCGGAGCCCTGCTTCTCTACACTGGCATGCTGATGAGATCGTCACATGCCACA	1809
Db	1140	GACCCGGAGCCCTGCTTCTCTACACTGGCATGCTGATGAGATCGTCACATGCCACA	1199
Qy	1810	TTGGCTCTTCCACATCTGGTTGACCTGTCATGATGGGCTCGCTCGATCTCCCTCAGTC	1869
Db	1200	TTGGCTCTTCCACATCTGGTTGACCTGTCATGATGGGCTCGCTCGATCTCCCTCAGTC	1259
Qy	1870	CCAAATTCAGTAGCCAAAGTGTTCCTCGACAGGCGTGTATGTTGCTGGCTGCCCAAGG	1929
Db	1260	CCAAATTCAGTAGCCAAAGTGTTCCTCGACAGGCGTGTATGTTGCTGGCTGCCCAAGG	1317
Qy	1930	GACACTCTCGCAGAGCCATTTTGGGTAAAGAACACTTACAAAGAGGCAATTCATCTGTT	1989
Db	1318	GACACTCTCGCAGAGCCATTTTGGGTAAAGAACACTTACAAAGAGGCAATTCATCTGTT	1377
Qy	1990	GTCTGAGGCTCAGAGCCCTTTTGATAGGCTTCTCATGTCATTCATAAAGACATTCGAAGCC	2049
Db	1378	GTCTGAGGCTCAGAGCCCTTTTGATAGGCTTCTCATGTCATTCATAAAGACATTCGAAGCC	1436
Qy	2050	AAGATGCTCCAACTGCAAATATACCAACCTTCTCTGAATTAATATTTTGCCTTATATAT	2109
Db	1437	AAGATGCTCCAACTGCAAATATACCAACCTTCTCTGAATTAATATTTTGCCTTATATAT	1496
Qy	2110	TCTTTTCTTTTTTCTAAAGAAATGGCTCTGAATAGATGCAACATTTCCATCTGAACTG	2169
Db	1497	TCTTTTCTTTTTTCTAAAGAAATGGCTCTGAATAGATGCAACATTTCCATCTGAACTG	1554
Qy	2170	GATGCATATCATTTAGCCAAATCCAGTAATTTATTAATTAATCTATACATAATATGTT	2229
Db	1555	GATGCATATCATTTAGCCAAATCCAGTAATTTATTAATTAATCTATACATAATATGTT	1614
Qy	2230	CCTCAGCATAGAGCTATGATTCATTAATTAAGAGTGGAGTCAAAAACGCTAAATGCAATG	2289
Db	1615	CCTCAGCATAGAGCTATGATTCATTAATTAAGAGTGGAGTCAAAAACGCTAAATGCAATG	1674
Qy	2290	TTTGTGTGATATTTTCATTAACAAACCTTAATTTGCTTTGTTAAATAAGTCAAGTGGAT	2349
Db	1675	TTTGTGTGATATTTTCATTAACAAACCTTAATTTGCTTTGTTAAATAAGTCAAGTGGAT	1733
Qy	2350	CTTGGAGTGGGATTTCTTGGTAAATTTCTTGCACCTTGAATGTCTCATGATTCATATGA	2409
Db	1734	CTTGGAGTGGGATTTCTTGGTAAATTTCTTGCACCTTGAATGTCTCATGATTCATATGA	1793
Qy	2410	AATCGCTTTGACATATCTTTTAGACAGAAAAAGTAGTGTGAGGGGGAAATTTATGAG	2469
Db	1794	AATCGCTTTGACATATCTTTTAGACAGAAAAAGTAGTGTGAGGGGGAAATTTATGAG	1853
Qy	2470	CTTGTGTGACTTTAGGGAGTAGCT	2493
Db	1854	C-TGTGTGACTTTAGGGAGTAGCT	1876

XX	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	Schlegel R, Endege WO, Monahan JE;
PI	WI; 2001-662795/76.
XX	Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.
PT	Claim 1; Page 5416-5417; 11750pp; English.
XX	The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX	Sequence 2199 BP; 597 A; 488 C; 516 G; 592 T; 0 U; 6 Other;
QY	Query Match 61.4%; Score 1579.6; DB 5; Length 2199;
DB	Best Local Similarity 97.1%; Pred. No. 2.5e-306;
QY	Matches 1748; Conservative 0; Mismatches 34; Indels 18; Gaps 13;
QY	695 GTTTTGGAGGAGTGACCTCCTTTATTCCTGAAGCCCGGCAGTGCTGTGACTGGTGCAC 754
DB	187 GCTTGTGCACAGTGACCTCCTTTATTCTGAAGCCCGGCAGTGCTGTGACTGGTGCAC 246
QY	755 CAGTCCCCAAACAGATCCAGAAGGAAGTTCAGGTTGTGCCATCGAAGGACTGGAT 814
DB	247 CAGTCCCCAAACAGATCCAGAAGGAAGTTCAGGTTGTGCCATCGAAGGACTGGAT 306
QY	815 GCCACGGGTAAACCAACCGTAGCCAGTCAGTGGCAGATTCACTTAAGGCTGTCTCTTA 874
DB	307 GCCACGGGTAAACCAACCGTAGCCAGTCAGTGGCAGATTCACTTAAGGCTGTCTCTTA 366
QY	875 AAGTCAACCACCTCTTGATTTGGCCAGTGAGGAGAATCTTTGATGATGAACCAACTATC 934
DB	367 AAGTCAACCACCTCTTGATTTGGCCAGTGAGGAGAATCTTTGATGATGAACCAACTATC 426
QY	935 ATTAGAAGAGCTTTTACTCTTTGGGGAATTAATTGTGGGCTCCGAATAAGCTTAAGAA 994
DB	427 ATTAGAAGAGCTTTTACTCTTTGGGCAATTAATTGTGGGCTCCGAATAAGCTTAAGAA 486
QY	995 TCTGCCAATCTCTGTGATTTAGACAGGACTGGCACAGCAGCCGACCATATGCCATA 1054
DB	487 TCTGCCAATCTCTGTGATTTAGACAGGACTGGCACAGCAGCCGACCATATGCCATA 546
QY	1055 GCACCTGAGGTGAGTGGGGGTCTCCAGCACCTGCCCCAGCCCATCACCTGTGTACAG 1114
DB	547 GCACCTGAGGTGAGTGGGGGTCTCCAGCACCTGCCCCAGCCCATCACCTGTGTACAG 606
QY	1115 TGCCACAGGAGCTGTCTAAAACCTGACCTTATCTGTCTGCTCAGTGTAGTCTCAGGAG 1174
DB	607 TGCCACAGGAGCTGTCTAAAACCTGACCTTATCTGTCTCAGTGTAGTCTCAGGAG 666
QY	1175 AGTTTCAGAGCTTCAGGCGCGGGGCATGGAGAGACCCAGGAAGAGCAACTTCGAG 1234
DB	667 AGTTTCAGAGCTTCAGGCGCGGGGCATGGAGAGACCCAGGAAGAGCAACTTCGAG 726
QY	1235 GCCAACAGTGTGTTCCTCAAAAGGTAGAAATGTCTTACAGCGGATGAGGAATCTCTGGC 1294
DB	727 GCCAACAGTGTGTTCCTCAAAAGGTAGAAATGTCTTACAGCGGATGAGGAATCTCTGGC 786
QY	1295 TGCCATGTGGTGTGATGCCACCCCTCCAGAGAAAAGGTCCTGCAGACAGTATTAGCCCTA 1354

787	TGCCATGTGGTGTGATGCCACGCCCTCCAGAGAAAAGGTCCTGCAGACGGTATTAAAGCCTA	846
Db		
QY	1355 ATCCAGAAATAGTTTTAGTGAACCGTAGTTACTCTGGCCAGGTTGCCACGCTTAACTAGATTT	1414
QY		
Db	847 ATCCAGAAATAGTTTTAGTGAACCGTAGTTACTCTGGCCAGGTTGCCACGCTTAACTAGATTT	906
Db		
QY	1415 AGATGTGTGTTTGAACAATCTCATCCACCATTTGTTATGCAAGTGTGCCAAATTTCTGTT	1474
QY		
Db	907 AGATGTGTGTTTGAACAATCTCATCCACCATTTGTTATGCAAGTGTGCCAAATTTCTGTT	966
Db		
QY	1475 CTACAAGCAATGTTGTGGCGAAGAACTGGAGACCAAGGCACTTAAATTTTACTTCAGCCA	1534
QY		
Db	967 CTACAAGCAATGTTGTGGCGAAGAACTGGAGACCAAGGCACTTAAATTTTACTTCAGCCA	1026
Db		
QY	1535 TGTATCCCCTCTCTGACTGATCGAGCCCGCTCATCAAAAGGTCCTCTCATCATGTTCCAG	1594
QY		
Db	1027 TGTATCCCCTCTCTGACTGATCGAGCCCGCTCATCAAAAGGTCCTCTCATCATGTTCCAG	1086
Db		
QY	1595 TGAGAGGCCAGCAATGCTTCTCTGGCATAGTAAACATTTCTTGSAACATATGTTT	1654
QY		
Db	1087 TGAGAGGCCAGCAATGCTTCTCTGGCATAGTAAACATTTCTTGSAACATATGTTT	1146
Db		
QY	1655 CACTTAATCACTACCAAAATPATCTGGAAGACCTGCTTACTCAGACAGACCAAGGTGTACA	1714
QY		
Db	1147 CACTTAATCACTACCAAAATPATCTGGAAGACCTGCTTACTCAGACAGACCAAGGTGTACA	1206
Db		
QY	1715 GAAGCAGCAGACAAGATCTTCCAGATCAGCAGSGAGACCCCGAGGCTCTGCTTCCTCTA	1774
QY		
Db	1207 GAAGCAGCAGACAAGATCTTCCAGATCAGCAGSGAGACCCCGAGGCTCTGCTTCCTCTA	1266
Db		
QY	1775 CACTGCGATCTGATCAGATCGTGACATGCCACATTTGGCTTCTTCCACATCTGCTGTGCA	1834
QY		
Db	1267 CACTGCGATCTGATCAGATCGTGACATGCCACATTTGGCTTCTTCCACATCTGCTGTGCA	1326
Db		
QY	1835 CTGCTCATGATGGGCTCGCTGCATCTCCCTCAGTCCCAAATCTCTAGTAGCAGCAAGTCTCC	1894
QY		
Db	1327 CTGCTCATGATGGGCTCGCTGCATCTCCCTCAGTCCCAAATCTCTAGTAGCAGCAAGTCTCC	1385
Db		
QY	1895 TGACAGGCTGTCTATGTGCTCTGGCTGCCAAGGACACTCCTCGACAGACCAATTTTGG	1954
QY		
Db	1386 TGACAGGCTGTCTATGTGCTCTGGCTGCCAAGGACACTCCTCGACAGACCAATTTTGG	1444
Db		
QY	1955 GTAAGGAACAATTACAAAAGGCAATGATCTTGTCTGAGGCTCAGAGCCCTTTTGTAT	2014
QY		
Db	1445 GTAAGGAACAATTACAAAAGGCAATGATCTTGTCTGAGGCTCAGAGCCCTTTTGTAT	1504
Db		
QY	2015 AGGCTTCTGATGTCATTAATAAGACATTAAGACCAAGATGCTCCAAATGATATACC	2074
QY		
Db	1505 AGGCTTCTGATGTCATTAATAAGACATTAAGACCAAGATGCTCCAAATGATATACC	1563
Db		
QY	2075 AACCTTCTGAAATTAATTTTGTGTTTATTAATTTTCTTTTCTTTTCTTTTCTTTTCTTT	2134
QY		
Db	1564 AA-CTTCTCTGAAATTAAT-TTTGCTTATTTATTTTCTTTTCTTTTCTTTTCTTTTCTTT	1619
Db		
QY	2135 GCTCTGAATAGAAATGACATTTTCCATCTCAACTGGAATGATATCATTTAGCCAAATCCAG	2194
QY		
Db	1620 GCTCTGAATAGAAATGACATTTTCCAT-TGAACTGGAATGATTTTCAATTTAGCCAAATCCAG	1678
Db		
QY	2195 TAATTTATTTATTAATCTATATATAAATATGTTTCTCAGCATAGAGCTATGATTTTCAAT	2254
QY		
Db	1679 TAATTTATTTATTAATCTATATATAAATATGTTTCTCAGCATAGAGCTATGATTTTCAAT	1738

Db 1851 TTATCTTGCACTTGAAATGCTCTCTGATTACATATGAAATCGC-TTGACATATCTTTGGAC 1909

QY 2434 AGAAAAAGTAGCTGAGTGGGGGAAATATAGAGCTTGTGTGACCTTTTAGGGAGTAGCT 2493

Db 1910 AGAAAAAAGTAGTGAAGTGGGGGAAATATAGAGCTTGTGTGACCTTTTAGGGAGTAGGT 1969

RESULT 15

ABV20963

ID ABV20963 standard; cDNA; 2199 BP.

XX AC ABV20963;

XX DT 13-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 20954.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WC200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US005171.

XX PR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX PX WPI; 2001-662795/76.

XX PT Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer.

XX PS Claim 1; Page 3446; 11750pp; English.

XX CC The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC determining whether prostate cancer has metastasized in a patient; (h)

CC assessing the aggressiveness or indolence of prostate cancer in a patient

CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX SQ Sequence 2199 BP; 597 A; 488 C; 516 G; 592 T; 0 U; 6 Other;

Query Match 61.4%; Score 1579.6; DB 5; Length 2199;

Best Local Similarity 97.1%; Pred. No. 2.5e-306;

Matches 1748; Conservative 0; Mismatches 34; Indels 18; Gaps 13;

QY 695 GTTTGGAGAGTGTACCTCTTTATTCTGAAGCCCGGCGAGTGTGACCTGTGCGAC 754

Db 187 GCTTGTGCACATGTACCTCTTTATTCCTGAAGCCCGGCGAGTGTGACCTGTGCGAC 246

QY 755 CAGTGCCTCAAAACAGATCCAGAAAGAAAGTTCCAGGTTGTTGCCATCGAAGGACTGGAT 814

Db 247 CAGTGCCTCAAAACAGATCCAGAAAGAAAGTTCCAGGTTGTTGCCATCGAAGGACTGGAT 306

815 GCCACGGGTAAAAACCAACCGGTGACCCAGTCACTGCGAGATTCACTTAAGGCTGTCTCTTA 874

Db 307 GCCACGGGTAAAAACCAACCGGTGACCCAGTCACTGCGAGATTCACTTAAGGCTGTCTCTTA 366

QY 875 AAGTCAACCCCTCTTGCATTTGCCAGTGGAGGAAGATCTTTGATGATGAACCACTATC 934

Db 367 AAGTCAACCCCTCTTGCATTTGCCAGTGGAGGAAGATCTTTGATGATGAACCACTATC 426

QY 935 ATTAGAAGAGCTTTTACTCTTTGGGCAATTATATTGTGGCTCCGAAATAGCTAAGAA 994

Db 427 ATTAGAAGAGCTTTTACTCTTTGGGCAATTATATTGTGGCTCCGAAATAGCTAAGAA 486

QY 995 TCTGCCAAATCTCTGTGATTGTAGACAGTACTGGACACAGCGGCACCTATGCCATA 1054

Db 487 TCTGCCAAATCTCTGTGATTGTAGACAGTACTGGACACAGCGGCACCTATGCCATA 546

QY 1055 GCCACTGAGGTGAGTGGGGTCTCCAGCACCTGCCCCCAGCCCATCACCTGTGTACCAG 1114

Db 547 GCCACTGAGGTGAGTGGGGTCTCCAGCACCTGCCCCCAGCCCATCACCTGTGTACCAG 606

QY 1115 TGGCCAGAGGACCTGCTCAAAACCTGACCTTATCTCTGCTGCTACTGCTGAGTCTCTGAGGAG 1174

Db 607 TGGCCAGAGGACCTGCTCAAAACCTGACCTTATCTCTGCTGCTACTGCTGAGTCTCTGAGGAG 666

QY 1175 AGTTGCGAGGCTGACGGCGGGCATGAGAGAACACGAGGAGAAAGCAGAACTTGAG 1234

Db 667 AGTTGCGAGGCTGACGGCGGGCATGAGAGAACACGAGGAGAAAGCAGAACTTGAG 726

QY 1235 GCCACAGTGTGTTGCTCAAAAGGTAGAAATGTTCTTACACGCGGATGAGATCTCTGCC 1294

Db 727 GCCACAGTGTGTTGCTCAAAAGGTAGAAATGTTCTTACACGCGGATGAGATCTCTGCC 786

QY 1295 TGCATGTGTTGATGCCAGCCCTCCAGAGAAAGGCTCTGCGACAGCTATTATTAAGCCTA 1354

Db 787 TGCATGTGTTGATGCCAGCCCTCCAGAGAAAGGCTCTGCGACAGCTATTATTAAGCCTA 846

QY 1355 ATCCGAATAGTTTAGTGAAACCGTAGTTACTCTGCGCAGGTGCCACGCTAACTAGATT 1414

Db 847 ATCCGAATAGTTTAGTGAAACCGTAGTTACTCTGCGCAGGTGCCACGCTAACTAGATT 906

QY 1415 AGATGTTGTTCAAAACATCTACATCCACATTTGTTATCCAGTGTTCCTCAAAATTTCTGTT 1474

Db 907 AGATGTTGTTCAAAACATCTACATCCACATTTGTTATCCAGTGTTCCTCAAAATTTCTGTT 966

QY 1475 CTACAAGCATGTTGTGTCGAGAAAACCTGGAGACAGGATCTTAATTTTACTTTCAGCCA 1534

Db 967 CTACAAGCATGTTGTGTCGAGAAAACCTGGAGACAGGATCTTAATTTTACTTTCAGCCA 1026

QY 1535 TCGTACCTCTTCTGACTGATGAGCCCGTCATCAAAAGGTCCCTCTCATCATGTTCCAG 1594

Db 1027 TCGTACCTCTTCTGACTGATGAGCCCGTCATCAAAAGGTCCCTCTCATCATGTTCCAG 1086

QY 1595 TGAGAGCCAGGATGCTTCTTCTGCGATAGTAACATTTCTTGGAAACATATGTTT 1654

Db 1087 TGAGAGCCAGGATGCTTCTTCTGCGATAGTAACATTTCTTGGAAACATATGTTT 1146

QY 1655 CACTTAATCACTACCAAAATATCTGGAAGACCTGTTACTCTCAGACAGCAGGTTGATA 1714

Db 1147 CACTTAATCACTACCAAAATATCTGGAAGACCTGTTACTCTCAGACAGCAGGTTGATA 1206

QY 1715 GAAGCAGCAGCAAGATCTTCCAGATCAGAGGAGACCCCGAGGCTTCTGCTCTCTTA 1774

Db 1207 GAAGCAGCAGCAAGATCTTCCAGATCAGAGGAGACCCCGAGGCTTCTGCTCTCTTA 1266

QY 1775 CACTGGCATGCTGATGAGATCGTGCATGCCACATTTGGCTTCTTCCACATCTGTTGCA 1834

Db 1267 CACTGGCATGCTGATGAGATCGTGCATGCCACATTTGGCTTCTTCCACATCTGTTGCA 1326

QY 1835 CTCGTGATGATGGGCTCGCTGCATCTCCTCAGTCCCAAAATTTCTAGTAGCCAAAGTGTCC 1894

Db 1327 CTCGTGATGATGGGCTCGCTGCATCTCCTCAGTCCCAAAATTTCTAGTAGCCAAAGTGTCC 1385

Sun Oct 24 15:19:27 2004

us-10-681-223-1.rng

QY	1895	TGCAGAGGCTGTCTATGTGTCTGTGCTGCCAAGGACACTCTCTGCAGAGCCATTTTGG	1954
Db	1386	TGCAGAGGCTGTCTATGTGTCTGTGCTGCCAA-GGACACTCTGCAGAGCCATTTTGG	1444
QY	1955	GTAAGGACACTTACAAAGGCAATTGATCTTGTCTGTAGGCTCAGAGCCCTTTTGAT	2014
Db	1445	GTAAGGACACTTACAAAGGCAATTGATCTTGTCTGTAGGCTCAGAGCCCTTTTGAT	1504
QY	2015	AGGCTTCTCATGTCAATCATTAAGACATTCAAGCCAAAGATGTCCTCAACTGCAAAATATACC	2074
Db	1505	AGGCTTCTGA-GTCATATATAAGACATTCAAGCCAAAGATGTCCTCAACTGCAAAATATACC	1563
QY	2075	AACCTTCTCTGAATTATATTTTGGCTTATTTATATTTCTTTCTTTTCTTAAAGATTG	2134
Db	1564	AA-CTTCTCTGAATTATA-TTTGGCTTATTTATATTTCTTTCTTTTCTTAAAGTA--TG	1619
QY	2135	GCTCTGAATAGATGACATTTTCCATCTGAAGTGGATGCATATCATTTAGCCAAATCCAG	2194
Db	1620	GCTCTGAATAGATGACATTTTCCAT-TGAAGTGGATGCATTTCAATTTAGCCAAATCCAG	1678
QY	2195	TAAATTTATTTATATTAATCTATACATAATATGTTTCTCTAGCATAGGAGCTATGATTCAT	2254
Db	1679	TAAATTTATTTATTAATCTATACATAATATGTTTCTCTAGCATAGGAGCTATGATTCAT	1738
QY	2255	TAAATTAAGTGAGTCAAAACGCTAAATGCAATGTTTGTGTGTTATTTTCATTACACAA	2314
Db	1739	TAAATTAAGTGAGTCAAAACGCTAAATGCAATGTTTGTGTGTTATTTTCATTACACAA	1794
QY	2315	ACTTAATTTCTGTTGTTAAATAAGTTCAAAGTGGATCTTGGAGTGGGATTTCTT-GGTAAA	2373
Db	1795	A--TTAATTTCTGTTGTTAAATAAGTTCAAAGTGGATCTTGGAGTGGGATTTCTTGGGTAAA	1850
QY	2374	TTATCTTGCACTTGAATGTCTCATGATTACATATGAATCGCTTTGACATATCTTTAGAC	2433
Db	1851	TTATCTTGCACTTGAATGTCTCTGATTACATATGAATCGC-ITGACATATCTTTGGAC	1909
QY	2434	AGAAAAAGTAGCTGAGTGAGGGGGAATTTATAGAGCTTGTGACCTTTAGGGAGTAGCT	2493
Db	1910	AGAAAAAGTAGCTGAGTGAGGGGGAATTTATAGAGCTTGTGACCTTTAGGGAGTAGGT	1969

Search completed: October 22, 2004, 15:07:42
Job time : 1235 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 22, 2004, 08:04:13 ; Search time 10840 Seconds

(without alignments)
11216.035 Million cell updates/sec

Title: US-10-681-223-1

Perfect score: 2571

Sequence: 1 cggggccgagggcgctgg.....aggaggggcttttgcaccc 2571

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sv.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2571	100.0	2571	6	AR258675	AR258675 Sequence
2	2571	100.0	2571	6	AR438694	AR438694 Sequence
3	1579.6	61.4	2199	6	CQ489087	CQ489087 Sequence
4	1579.6	61.4	2199	6	CQ490198	CQ490198 Sequence
5	1579.6	61.4	2199	6	CQ494933	CQ494933 Sequence
6	1579.6	61.4	2199	6	CQ496057	CQ496057 Sequence
7	1459.6	56.8	1535	6	AX814311	AX814311 Sequence
8	1336.6	52.0	1343	6	AX814308	AX814308 Sequence
9	1193.8	46.4	1720	9	BC016969	BC016969 Homo sapi
10	1141.4	44.4	11808	9	AC017076	AC017076 Homo sapi
11	1130.4	44.0	20966	6	AR258676	AR258676 Sequence
12	1130.4	44.0	20966	6	AR438695	AR438695 Sequence
13	1065.2	41.4	1215	6	CQ733692	CQ733692 Sequence
14	889.2	34.6	1609	9	AK127983	AK127983 Homo sapi
15	873	34.0	3028	10	BC057565	BC057565 Mus muscu
16	797.2	31.0	1188	6	CQ777457	CQ777457 Sequence
17	690.2	26.8	757	6	AX814328	AX814328 Sequence
18	674.8	26.2	759	6	AX814327	AX814327 Sequence
19	646.6	25.1	833	6	AX814312	AX814312 Sequence

20	621.8	24.2	844	6	AX814319	AX814319 Sequence
21	600.4	23.4	972	6	AX814313	AX814313 Sequence
22	569	22.1	728	6	AX814321	AX814321 Sequence
23	562	21.9	700	6	AX814315	AX814315 Sequence
24	562	21.9	700	6	AX814322	AX814322 Sequence
25	534	20.8	702	6	AX814323	AX814323 Sequence
26	533.2	20.7	795	6	AX814316	AX814316 Sequence
27	524.2	20.4	669	6	AX814329	AX814329 Sequence
28	506	19.7	1008	6	AX814318	AX814318 Sequence
29	493.8	19.2	928	6	AX814326	AX814326 Sequence
30	482.6	18.8	556	6	AX814330	AX814330 Sequence
31	480.4	18.7	674	6	AX814317	AX814317 Sequence
32	463.6	18.0	3292	6	AX306082	AX306082 Sequence
33	463.6	18.0	3292	10	MUSTHYKIN	L329773 Mouse thymi
34	456.6	17.8	633	6	AX814314	AX814314 Sequence
35	453.4	17.6	531	6	CQ776571	CQ776571 Sequence
36	411.4	16.0	454	6	CQ501652	CQ501652 Sequence
37	411.4	16.0	454	6	CQ508973	CQ508973 Sequence
38	411.4	16.0	454	6	CQ510575	CQ510575 Sequence
39	411.2	16.0	418	6	CQ480481	CQ480481 Sequence
40	399	15.5	417	6	CQ478818	CQ478818 Sequence
41	390.8	15.2	454	6	CQ471313	CQ471313 Sequence
42	382.8	14.9	432	6	CQ500010	CQ500010 Sequence
43	360.6	14.0	443	11	G23005	G23005 human STS W
44	360	14.0	226460	2	AC117652	AC117652 Mus muscu
45	360	14.0	309317	2	AC132190	AC132190 Mus muscu

ALIGNMENTS

RESULT 1
AR258675
LOCUS AR258675 2571 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 1 from patent US 6489153.
ACCESSION AR258675
VERSION AR258675.1 GI:27309061
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2571)
AUTHORS Wei, M.-H., Ketchum, K.A., Beasley, E.M. and DiFrancesco, V.
TITLE Isolated human kinase proteins, nucleic acid molecules encoding human kinase proteins, and uses thereof
JOURNAL Patent: US 6489153-A 1 03-DEC-2002;
FEATURES
Location/Qualifiers
source
1..2571
/organism="unknown"
/mol_type="genomic DNA"

Query Match	100.0%	Score 2571;	DB 6;	Length 2571;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2571;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	CGGGGCGAGGGCGGCGTGCCTGAGCGGCCCATGGGCTTTGCGCGCGCGCTCTTGGCGCG 60		
DB	1	CGGGGCGAGGGCGGCGTGCCTGAGCGGCCCATGGGCTTTGCGCGCGCGCTCTTGGCGCG 60		
QY	61	GCACCTGTGCGGGCGGCGTGCCTGCGGGCGGCGGCTGCGGTGGGGCCCATGGCTCCGCC 120		
DB	61	GCACCTGTGCGGGCGGCGTGCCTGCGGGCGGCGGCTGCGGTGGGGCCCATGGCTCCGCC 120		
QY	121	GTGCGCGCTTGGCTTGGAGCTTCCGACCTGACCTGGCTCACTTCGCCCTAGGGCGCGA 180		
DB	121	GTGCGCGCTTGGCTTGGAGCTTCCGACCTGACCTGGCTCACTTCGCCCTAGGGCGCGA 180		
QY	181	CGCCCCCGGCGACGACAGCGCCCCCGACCCCGCTGGCGGCGCTGTTGGGCCCCCGGA 240		
DB	181	CGCCCCCGGCGACGACAGCGCCCCCGACCCCGCTGGCGGCGCTGTTGGGCCCCCGGA 240		
QY	241	GCGCAGCTACTCGCTGTGGTGGTCCCGTGACCCCGGACGCCGCGCTGGGGGCCCGGGTCCG 300		

Db	241	GGCGAGCTACTCGCTGTGGTGCCCGTGACCCCGGACGCGCGTGCGGGCGCGGGTCGG	300	1321	CAGAGAAAAGGTCCTGCAGACAGTATTAAAGCCCTAATCCAGAATAGTTTGTAGTGAACCGTA	1380
Qy	301	GGCGGGGGCTGCACAGGCGCTGTGACACAGCTGCGCCGCGGCGCCCTTCCAGCGGTG	360	1381	GTTACTCTGGCCAGGTGCCACGTCCTAACTAGATTAGATGTTTGTGAAACATCTACATCC	1440
Db	301	GGCGGGGGGGCTGCACAGGCGCTGTGACACAGCTGCGCCGCGGCGCCCTTCCAGCGGTG	360	1381	GTTACTCTGGCCAGGTGCCACGTCCTAACTAGATTAGATGTTTGTGAAACATCTACATCC	1440
Qy	361	CCAGCTGCTCAGGCTCTGTGTTACTGCCCGGGCGGCGAGGCGGGCGGCGACAGCAAGG	420	1441	ACCAATTGTTATGCAAGTGTCCCAAATTTCTGTCTACAAGCATGTGTGTGTCGACAAAA	1500
Db	361	CCAGCTGCTCAGGCTCTGTGTTACTGCCCGGGCGGCGAGGCGGGCGGCGACAGCAAGG	420	1441	ACCAATTGTTATGCAAGTGTCCCAAATTTCTGTCTACAAGCATGTGTGTGTCGACAAAA	1500
Qy	421	CTTCTGCTGCGGACCCCTGTGATGACCTGTGACCCCGGACGCGCTGCTCGAGTGCT	480	1501	CTGGAGACCAGGCATCTTAATTTTACTTTCAGCCATCGTACCCCTTCTCTGACTGATGGAC	1560
Db	421	CTTCTGCTGCGGACCCCTGTGATGACCTGTGACCCCGGACGCGCTGCTCGAGTGCT	480	1501	CTGGAGACCAGGCATCTTAATTTTACTTTCAGCCATCGTACCCCTTCTCTGACTGATGGAC	1560
Qy	481	GGCGCTCTCCAGAGGACACCGCCGCACTTGGGCGAGTTCCAGGCGCCACCCGCGCGG	540	1561	CGTCACTACAAAAGTCCCTCTCATCATGTTCAGTGTCAGTGAGGCGCAGGATTTCTTCTTC	1620
Db	481	GGCGCTCTCCAGAGGACACCGCCGCACTTGGGCGAGTTCCAGGCGCCACCCGCGCGG	540	1561	CGTCACTACAAAAGTCCCTCTCATCATGTTCAGTGTCAGTGAGGCGCAGGATTTCTTCTTC	1620
Qy	541	CCAGCTGTGCGACGCGCTCTGGAGGTGCAAGACGCGAGCGCGCTGCGAGTGGGCTGGCG	600	1621	TGGCATAGTAAACATTTTCTTGGAACATATGTTTCACTTAATCACCTACCAAAATATCTGA	1680
Db	541	CCAGCTGTGCGACGCGCTCTGGAGGTGCAAGACGCGAGCGCGCTGCGAGTGGGCTGGCG	600	1621	TGGCATAGTAAACATTTTCTTGGAACATATGTTTCACTTAATCACCTACCAAAATATCTGA	1680
Qy	601	ACAGTGTGCTCCCGTCCCGGACCCCGCTGACCCGCTGTCACCGGTGTCGACACTTGC	660	1681	AGACCTGTCTTACTCAGACAGCACCGAGGTGTAAGAAGCAGCAGACAGATCTTCCAGAT	1740
Db	601	ACAGTGTGCTCCCGTCCCGGACCCCGCTGACCCGCTGTCACCGGTGTCGACACTTGC	660	1681	AGACCTGTCTTACTCAGACAGCACCGAGGTGTAAGAAGCAGCAGACAGATCTTCCAGAT	1740
Qy	661	CGTGTGCTTCCCGACCGGACCGCGCCGCGCGCTTGGAGGAGTGTACCTCTTTTAT	720	1741	CAGCAGGGAGACCCCGGAGGCTCTGCTTCTTCTTACACTGCGCATGCGCATGCGATCGTGAC	1800
Db	661	CGTGTGCTTCCCGACCGGACCGCGCCGCGCGCTTGGAGGAGTGTACCTCTTTTAT	720	1741	CAGCAGGGAGACCCCGGAGGCTCTGCTTCTTCTTACACTGCGCATGCGCATGCGATCGTGAC	1800
Qy	721	TCCTGAAGCCCGGCGAGTGTTCACCTGTGTCGACGAGTGCACCAAAACAGATCCAGAAAG	780	1801	ATGCCACACATGTCCTTCCACATCTGGTTGTCACCTCGTCAATGATGGGCTCGCTGCACT	1860
Db	721	TCCTGAAGCCCGGCGAGTGTTCACCTGTGTCGACGAGTGCACCAAAACAGATCCAGAAAG	780	1801	ATGCCACACATGTCCTTCCACATCTGGTTGTCACCTCGTCAATGATGGGCTCGCTGCACT	1860
Qy	781	AAAGTTCAGGTTGTCGACAGAGCTGGAATGCAAGGTAACCAACAGGTCAGCCCA	840	1861	CCCTCAGTCCCAAATTTCTAGTAGCAAGTGTTCCTGCGAGAGGCTGTCTATGTGTCTGGC	1920
Db	781	AAAGTTCAGGTTGTCGACAGAGCTGGAATGCAAGGTAACCAACAGGTCAGCCCA	840	1861	CCCTCAGTCCCAAATTTCTAGTAGCAAGTGTTCCTGCGAGAGGCTGTCTATGTGTCTGGC	1920
Qy	841	GTCAAGTGGCAGATTCACCTAAGGCTGTCCTTAAAGTCAACCCCTCTTGCATGGCCA	900	1921	TGCCCCAAGGACACTCTCGCAGAGCCATTTTGGGTAAAGAACACTTACAAAGAGGCAT	1980
Db	841	GTCAAGTGGCAGATTCACCTAAGGCTGTCCTTAAAGTCAACCCCTCTTGCATGGCCA	900	1921	TGCCCCAAGGACACTCTCGCAGAGCCATTTTGGGTAAAGAACACTTACAAAGAGGCAT	1980
Qy	901	GTGAGGAAGATCTTTGATGATGAACCAACTATCATTTAGAGAGCTTTTACTTTTGGG	960	1981	TGATCTTGTCTGTGAGGCTCAGAGCCCTTTTGTATAGGCTTCTGATGTCAATTAAGAC	2040
Db	901	GTGAGGAAGATCTTTGATGATGAACCAACTATCATTTAGAGAGCTTTTACTTTTGGG	960	1981	TGATCTTGTCTGTGAGGCTCAGAGCCCTTTTGTATAGGCTTCTGATGTCAATTAAGAC	2040
Qy	961	CAATTATATTGGGCTCCGAAATAGCTAAGATCTGCCAAATCTCTGTGATTGTAGA	1020	2041	ATTCAAGCCCAAGATGCTCCAACTGCAATATATACCAACTTCTCTGAAATTAATTTTGGTT	2100
Db	961	CAATTATATTGGGCTCCGAAATAGCTAAGATCTGCCAAATCTCTGTGATTGTAGA	1020	2041	ATTCAAGCCCAAGATGCTCCAACTGCAATATATACCAACTTCTCTGAAATTAATTTTGGTT	2100
Qy	1021	CAGTACTGGCAGACGCGCCACTATGCCATAGCCACTGAGTGAGTGGGGTCTCCA	1080	2101	ATTTATATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	2160
Db	1021	CAGTACTGGCAGACGCGCCACTATGCCATAGCCACTGAGTGAGTGGGGTCTCCA	1080	2101	ATTTATATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	2160
Qy	1081	GCACTGCCCGCCAGCCCATCACTGTGTGACAGTGGCGAGAGCACTGTCTCAAACTGA	1140	2161	TCCTGAACTGGATGCATATCAATTTAGCCAAATCCAGTAATTTATTTATTTATTTATAT	2220
Db	1081	GCACTGCCCGCCAGCCCATCACTGTGTGACAGTGGCGAGAGCACTGTCTCAAACTGA	1140	2161	TCCTGAACTGGATGCATATCAATTTAGCCAAATCCAGTAATTTATTTATTTATTTATAT	2220
Qy	1141	CCCTATCTCTGCTGCTGAGTCTGAGGAGGTTGAGGAGGTTGAGGAGGTTGAGGAGG	1200	2221	AATATGTTTCTCAGCAGTAGAGCTATGATTCATTAATTAAGTGGAGTCAAAACGCTA	2280
Db	1141	CCCTATCTCTGCTGCTGAGTCTGAGGAGGTTGAGGAGGTTGAGGAGGTTGAGGAGG	1200	2221	AATATGTTTCTCAGCAGTAGAGCTATGATTCATTAATTAAGTGGAGTCAAAACGCTA	2280
Qy	1201	CATGAGAGAACCCAGGAGAGCAGAACTTTCAGGCCAACAGTGTGTTCCTCAAAAGT	1260	2281	AATGCAATGTTTGTGTGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	2340
Db	1201	CATGAGAGAACCCAGGAGAGCAGAACTTTCAGGCCAACAGTGTGTTCCTCAAAAGT	1260	2281	AATGCAATGTTTGTGTGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	2340
Qy	1261	AGAAATGCTTACAGCGGATGGAGATCTTGGCTGCGCATGTGTGTGATGCGAGCCCTC	1320	2341	CAAGTGGATCTTGGAGTGGGATTTCTTGGTAAATTAATCTTGCATCTGAAATGCTCATGAT	2400
Db	1261	AGAAATGCTTACAGCGGATGGAGATCTTGGCTGCGCATGTGTGTGATGCGAGCCCTC	1320	2341	CAAGTGGATCTTGGAGTGGGATTTCTTGGTAAATTAATCTTGCATCTGAAATGCTCATGAT	2400
Qy	1321	CAGAGAAAAGTCTCTGCAGACAGTATTAAAGCCCTAATCCAGAATAGTTTGTAGTGAACCGTA	1380	2401	TACATATGAATTCGCTTTGACATATCTTTAGACAGAAAAGTGTAGTGTGAGGGGAA	2460
Db	1321	CAGAGAAAAGTCTCTGCAGACAGTATTAAAGCCCTAATCCAGAATAGTTTGTAGTGAACCGTA	1380	2401	TACATATGAATTCGCTTTGACATATCTTTAGACAGAAAAGTGTAGTGTGAGGGGAA	2460

```
QY 2461 ATTATAGAGCTTCTGTGACTTTAGGAGTAGCTGTCTCTTATACACATCTCAAGCCCTG 2520
Db 2461 ATTATAGAGCTTCTGTGACTTTAGGAGTAGCTGTCTCTTATACACATCTCAAGCCCTG 2520
QY 2521 AAGCCTTGATGTCCTGCGAGCTGCGACTAAAGAGGGGGCTTTTGCACCC 2571
Db 2521 AAGCCTTGATGTCCTGCGAGCTGCGACTAAAGAGGGGGCTTTTGCACCC 2571

RESULT 2
AR438694
LOCUS AR438694 2571 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 1 from patent US 6664087.
ACCESSION AR438694
VERSION AR438694.1 GI:42663635
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2571)
AUTHORS Wei,M.-H., Ketchum,K.A., Beasley,E.M. and Difrancesco,V.
TITLE Isolated human kinase proteins
JOURNAL Patent: US 6664087-A 1 16-DEC-2003;
FEATURES
    Location/Qualifiers
        source
            1..2571
                /organism="unknown"
                /mol_type="genomic DNA"

ORIGIN
Query Match 100.0%; Score 2571; DB 6; Length 2571;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGGCGAGGGCGGCTGCTGAGCGCCCATGCGCTTCGCGCGCGGCTCCTCGCGG 60
Db 1 CGGGGCGAGGGCGGCGGCTGCTGAGCGCCCATGCGCTTCGCGCGCGGCTCCTCGCGG 60
QY 61 GCGACTGTCGGGCGGCTGCTGCGGCGGCGGCGGCTGCGCTGCGGCGGCGGCTCGG 120
Db 61 GCGACTGTCGGGCGGCGGCTGCTGCGGCGGCGGCGGCTGCGCTGCGGCGGCGGCTCGG 120
QY 121 GTGCGGCTTCTGCTGAGGCTTCCGACTGCACTGCACTGGCTGCTGCTGCGCCCTAGGCGCCGA 180
Db 121 GTGCGGCTTCTGCTGAGGCTTCCGACTGCACTGCACTGGCTGCTGCTGCGCCCTAGGCGCCGA 180
QY 181 CGCGCCCGCGAGCGAGAGCGCCCGGACCCCGGCTGCGGCGGCTGTTGGGCGCCCGGA 240
Db 181 CGCGCCCGCGAGCGAGAGCGCCCGGACCCCGGCTGCGGCGGCTGTTGGGCGCCCGGA 240
QY 241 GCGAGCTACTCGCTGTGCGTGGTGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGG 300
Db 241 GCGAGCTACTCGCTGTGCGTGGTGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGG 300
QY 301 GCGGCGGCGGCTGCAACGCGCTGCTGCAACGCGCTGCGCGCGGCGGCGGCGGCGG 360
Db 301 GCGGCGGCGGCTGCAACGCGCTGCTGCAACGCGCTGCGCGCGGCGGCGGCGGCGG 360
QY 361 CCAGTGTCTCAGGCTGCTCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 361 CCAGTGTCTCAGGCTGCTCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 421 CTTCTGCTGCGGACCCCTGATGATCCTGACACCGGCGAGGCTGCTGCTGCTGCTGCT 480
Db 421 CTTCTGCTGCGGACCCCTGATGATCCTGACACCGGCGAGGCTGCTGCTGCTGCTGCT 480
QY 481 GGGCGGCTGCCAGAGGACACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 481 GGGCGGCTGCCAGAGGACACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 CCAGTGTGCGAGCGGCTCTGGAGGCTGCAAGACGCGAGCGGCGGCTGCGAGTGGCTGCGC 600
Db 541 CCAGTGTGCGAGCGGCTCTGGAGGCTGCAAGACGCGAGCGGCGGCTGCGAGTGGCTGCGC 600
```

```
QY 601 AAGGTCGTGCGGTCCTCCCGAGCCCGCGCTGCAACCGGCTGTCACAGACTTGCCAGCTTC 660
Db 601 AAGGTCGTGCGGTCCTCCCGAGCCCGCGCTGCAACCGGCTGTCACAGACTTGCCAGCTTC 660
QY 661 CGTGGTCTTCCCGGACCGGAGACCGCGCGGCGGCTTTGGAGGAGTGTACCTCTTTAT 720
Db 661 CGTGGTCTTCCCGGACCGGAGACCGCGCGGCGGCTTTGGAGGAGTGTACCTCTTTAT 720
QY 721 TCCCTGAAGCCCGGCGAGTGTGCTTGAACCTGCTGACACAGTGCACCAACAGATCCAGAAAGG 780
Db 721 TCCCTGAAGCCCGGCGAGTGTGCTTGAACCTGCTGACACAGTGCACCAACAGATCCAGAAAGG 780
QY 781 AAGTTCCAGGTTTGGTGGCATCGAAGGACTGATGACACCGGCTGATGACACCGGCTGAC 840
Db 781 AAGTTCCAGGTTTGGTGGCATCGAAGGACTGATGACACCGGCTGATGACACCGGCTGAC 840
QY 841 GTCAAGTGGGAGATTCACCTTAAAGGCTGCTCTTAAAGTCAACACCTCTTGCATTCGCCA 900
Db 841 GTCAAGTGGGAGATTCACCTTAAAGGCTGCTCTTAAAGTCAACACCTCTTGCATTCGCCA 900
QY 901 GTGGAGAAAGATCTTTGATGATCAACCAACTATCATTAGAAGAGCTTTTACTCTTTGG 960
Db 901 GTGGAGAAAGATCTTTGATGATCAACCAACTATCATTAGAAGAGCTTTTACTCTTTGG 960
QY 961 CAATTATATTGTGGCTCCGAAATAGCTAAAGATCTGCAAAATCTCTGTGATTGTAGA 1020
Db 961 CAATTATATTGTGGCTCCGAAATAGCTAAAGATCTGCAAAATCTCTGTGATTGTAGA 1020
QY 1021 CAGGTACTGCGACAGCAGCGCCACCTATGCCATAGCAGCTGAGGTGAGTGGGGTCTCCA 1080
Db 1021 CAGGTACTGCGACAGCAGCGCCACCTATGCCATAGCAGCTGAGGTGAGTGGGGTCTCCA 1080
QY 1081 GCACCTGCGCCCGAGCCATCACCTGTGTACAGTGGCCAGAGGAGCTGCTCAAACTGA 1140
Db 1081 GCACCTGCGCCCGAGCCATCACCTGTGTACAGTGGCCAGAGGAGCTGCTCAAACTGA 1140
QY 1141 CTTATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db 1141 CTTATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
QY 1201 CATGAGAGAGACCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 CATGAGAGAGACCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 AGAAATGCTCTACAGCGGATGAGAAATCCTGGCTGCAATGCTGGTGTGATGCGAGCCCTC 1320
Db 1261 AGAAATGCTCTACAGCGGATGAGAAATCCTGGCTGCAATGCTGGTGTGATGCGAGCCCTC 1320
QY 1321 CAGAGAAAAGTCTCTGCGAGACAGTATTAAGCCTAATCCAGAAATAGTTTTAGTGAACCGTA 1380
Db 1321 CAGAGAAAAGTCTCTGCGAGACAGTATTAAGCCTAATCCAGAAATAGTTTTAGTGAACCGTA 1380
QY 1381 GTTACTCTGGCGAGTGCACCTGCTAACTAGATTAGATTGTTGTAACATCTACATCC 1440
Db 1381 GTTACTCTGGCGAGTGCACCTGCTAACTAGATTAGATTGTTGTAACATCTACATCC 1440
QY 1441 ACCATTTGTTATGAGTGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Db 1441 ACCATTTGTTATGAGTGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 CTGGAGACAGGAGTCTTTAATTTTACTCTGAGCCATCGTACCTCTTCTGATGTGGACC 1560
Db 1501 CTGGAGACAGGAGTCTTTAATTTTACTCTGAGCCATCGTACCTCTTCTGATGTGGACC 1560
QY 1561 CGTCATCAAGAGTCCCTCATCATGTTCCAGTGAAGCGGAGGAGTGTGTTCTTCTCC 1620
Db 1561 CGTCATCAAGAGTCCCTCATCATGTTCCAGTGAAGCGGAGGAGTGTGTTCTTCTCC 1620
QY 1621 TGGCATAGTAAACATTTTCTGGAAACATATGTTTCTCACTTAATCACTACCAATATCTGGA 1680
Db 1621 TGGCATAGTAAACATTTTCTGGAAACATATGTTTCTCACTTAATCACTACCAATATCTGGA 1680
```

QY	1681	AGACCTGCTTACTCAGACAGCACAGGCTGTACAGAGCAGACAGATCTTCAGAT	1740	AUTHORS	Schlegel, R., Endege, W.O. and Monahan, J.E.
Db	1681	AGACCTGCTTACTCAGACAGCACAGGCTGTACAGAGCAGACAGATCTTCAGAT	1740	TITLE	Genes differentially expressed in human prostate cancer and their use
QY	1741	CAGCAGGAGACCCGGGCTCTCTCTCTCAGACAGGCTGTACAGAGCAGATCTTCAGAT	1800	JOURNAL	Patent: WO 0160860-A 20954 23-AUG-2001;
Db	1741	CAGCAGGAGACCCGGGCTCTCTCTCTCAGACAGGCTGTACAGAGCAGATCTTCAGAT	1800	Millennium Predictive Medicine, Inc. (US)	
QY	1801	ATGCCACATCTGCTCTTCCACATCTGGTGTGACCTCGTCAATGAGGCTCGCTCATCT	1860	FEATURES	Location/Qualifiers
Db	1801	ATGCCACATCTGCTCTTCCACATCTGGTGTGACCTCGTCAATGAGGCTCGCTCATCT	1860	source	1. .2199
QY	1861	CCCTCAGTCCCAATTTCTAGTACCCAACTGCTCTCTCTCAGACAGGCTGTCTATGCTCTGGC	1920	ORIGIN	/organism="Homo sapiens"
Db	1861	CCCTCAGTCCCAATTTCTAGTACCCAACTGCTCTCTCTCAGACAGGCTGTCTATGCTCTGGC	1920	Query Match	/mol_type="unassigned DNA"
QY	1921	TGCCCAAGGACACTCTCTCAGAGCCATTTTGGTAAAGAACACTTACAAAGAGGCAT	1980	Best Local Similarity	/db_xref="taxon:9606"
Db	1921	TGCCCAAGGACACTCTCTCAGAGCCATTTTGGTAAAGAACACTTACAAAGAGGCAT	1980	Matches 1748; Conservative	0; Mismatches 34; Indels 18; Gaps 13;
QY	1981	TGATCTTGTCTGAGGCTCAGAGCCCTTTTGTAGAGCTTCTGATGTCATCTCAAGAC	2040		
Db	1981	TGATCTTGTCTGAGGCTCAGAGCCCTTTTGTAGAGCTTCTGATGTCATCTCAAGAC	2040		
QY	2041	ATTCAAGCCAAAGTCTCAGACAGGCTCTCTCTCTCAGACAGGCTGTCTATGCTCTGGC	2100		
Db	2041	ATTCAAGCCAAAGTCTCAGACAGGCTCTCTCTCTCAGACAGGCTGTCTATGCTCTGGC	2100		
QY	2101	ATTCAAGCCAAAGTCTCAGACAGGCTCTCTCTCTCAGACAGGCTGTCTATGCTCTGGC	2160		
Db	2101	ATTCAAGCCAAAGTCTCAGACAGGCTCTCTCTCTCAGACAGGCTGTCTATGCTCTGGC	2160		
QY	2161	TCTGAACTGGATGCATATCAATTTAGCCAAATCCAGTAATTTATTTATTTATTTATAT	2220		
Db	2161	TCTGAACTGGATGCATATCAATTTAGCCAAATCCAGTAATTTATTTATTTATTTATAT	2220		
QY	2221	AATATGTTCTCAGATAGGAGCTATGATTTCAATTAATTAATTAATTAATTAATTAAT	2280		
Db	2221	AATATGTTCTCAGATAGGAGCTATGATTTCAATTAATTAATTAATTAATTAATTAAT	2280		
QY	2281	AATGCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	2340		
Db	2281	AATGCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	2340		
QY	2341	CAAGTGGATCTGGAGTGGGATTTCTTGGTAAATTAATTTGCACTTGAATGCTCATGAT	2400		
Db	2341	CAAGTGGATCTGGAGTGGGATTTCTTGGTAAATTAATTTGCACTTGAATGCTCATGAT	2400		
QY	2401	TACATATGAAATCGCTTTGACATATCTTTAGCAGAAAGAGTAGCTGAGTGGGGAA	2460		
Db	2401	TACATATGAAATCGCTTTGACATATCTTTAGCAGAAAGAGTAGCTGAGTGGGGAA	2460		
QY	2461	ATTATAGAGCTGTGTGACTTTAGGAGTAGCTGTCTTTATACATATCACTCAAGCCCTG	2520		
Db	2461	ATTATAGAGCTGTGTGACTTTAGGAGTAGCTGTCTTTATACATATCACTCAAGCCCTG	2520		
QY	2521	AAGCCTTGCATGCTCTGAGGCTGCACTTAAAGAGGGGCTTTTGACCC	2571		
Db	2521	AAGCCTTGCATGCTCTGAGGCTGCACTTAAAGAGGGGCTTTTGACCC	2571		
RESULT 3				CQ489087	
LOCUS				Sequence 20954 from Patent WO0160860.	
DEFINITION				CQ489087	
ACCESSION				CQ489087.1	
VERSION				GI:41454706	
KEYWORDS				Homo sapiens (human)	
SOURCE				Homo sapiens	
ORGANISM				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
				Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE				1	

Db 1027 TCGTACCTCTCTGACTGATGACCCGTCATCAAAAGGTCCTCTCATCATGTTCCAG 1086
QY 1595 TGAGAGCCGAGGATTGCTTTCTCTGTCATAGTAAACATTTCTTGTGAACATATGTTT 1654
Db 1087 TGAGAGCCGAGGATTGCTTTCTCTGTCATAGTAAACATTTCTTGTGAACATATGTTT 1146
QY 1655 CACTTAATCACTACCAATATCTGGAGACCTGTCTTACTCAGACGACCCAGGTGTACA 1714
Db 1147 CACTTAATCACTACCAATATCTGGAGACCTGTCTTACTCAGACGACCCAGGTGTACA 1206
QY 1715 GAAGCAGACAGCAAGATCTTCCAGATCAGACGGAGACCCCGAGGCTCTGTTCTCTTA 1774
Db 1207 GAAGCAGACAGCAAGATCTTCCAGATCAGACGGAGACCCCGAGGCTCTGTTCTCTTA 1266
QY 1775 CACTGCACTGCTGATGAGATCGTGACATGCCACATTTGGCTTTCTCCACATCTGTTGCA 1834
Db 1267 CACTGCACTGCTGATGAGATCGTGACATGCCACATTTGGCTTTCTCCACATCTGTTGCA 1326
QY 1835 CTGCTCATGATGGCTCGCTGATCTCCCTCAGTCCCAATTTCTAGTAGCCAAAGTGTTC 1894
Db 1327 CTCGTCATGATGGCTCGCTGATCTCCCTCAGTCCCAATTTCTAGTAGCCAAAGTGTTC 1385
QY 1895 TGCAGAGGCTGTCTATGCTGCTGCTGCCCAAGGACACTCTCTGAGAGCCATTTTGG 1954
Db 1386 TGCAGAGGCTGTCTATGCTGCTGCCCAAGGACACTCTCTGAGAGCCATTTTGG 1444
QY 1955 GTAAGAACACATTACAAGAGGATTTGATCTTGCTGCTGAGGCTCAGAGCCCTTTTGTAT 2014
Db 1445 GTAAGAACACATTACAAGAGGATTTGATCTTGCTGCTGAGGCTCAGAGCCCTTTTGTAT 1504
QY 2015 AGGCTTCTGATCTCATTAAGACATTCAGCCAGGATGCTCCAACTGCAAAATATACC 2074
Db 1505 AGGCTTCTGA-GTCATATAAAGACATTCAGCCAGGATGCTCCAACTGCAAAATATACC 1563
QY 2075 AACCTTCTCTGAATATATTTGCTTATTTATTTATTTCTTTCTTTTCTTAAAGATG 2134
Db 1564 AA-CTTCTCTGAATATA-TTGTCTATTTATTTCTTTCTTTTCTTAAAGTA--TG 1619
QY 2135 GCTCTGAATAGATGACATTTTCCATCTGAATGCAATCATATCATTTAGGCAATCCAG 2194
Db 1620 GCTCTGAATAGATGACATTTTCCAT-TGAATGCAATGCAATTTTCTTTAGGCAATCCAG 1678
QY 2195 TAATTTATTTATTAATCTATACATATGCTTTCTCAGCATAGGACTATGTTTAT 2254
Db 1679 TAATTTATTTATTAATCTATACATATGCTTTCTCAGCATAGGACTATGTTTAT 1738
QY 2255 TAATTAAGAGTGGAGTCAAAAGCGTAAATGCAATGTTTGTGTATTTTCTTATACAA 2314
Db 1739 TAATTAAGAGTGGAGTCAAAA-GCTAAATGCAATGT--GTGTATTTTCTTATACAA 1794
QY 2315 ACTTAATTTGCTGTTTAAATTAAGTTCAAGTGGATCTTGGATGGGATTTCTT-GGTAAA 2373
Db 1795 A--TTAATTTGCTGTTTAAATTAAGTACA--GGGATCTTGGATGGGATTTCTTGGTAAA 1850
QY 2374 TTATCTGCATGTAATGCTCATGATTTACATATGAAATCGCTTTGACATATCTTTAGAC 2433
Db 1851 TTATCTTGCACTGTAATGCTCTGATTTACATATGAAATCGC-TTGACATATCTTTGGAC 1909
QY 2434 AGAAAAAGTAGCTGAGTGAGGGGGAATTTATAGAGCTTGTGACTTTAGGGAGTAGCT 2493
Db 1910 AGAAAAAGTAGCTGAGTGAGGGGGAATTTATAGAGCTTGTGACTTTAGGGAGTAGCT 1969

RESULT 4
LOCUS CQ490198 2199 bp DNA linear PAT 30-JAN-2004
DEFINITION Sequence 22065 from Patent WO0160860.
ACCESSION CQ490198
VERSION CQ490198.1 GI:41455817
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Schlegel,R., Endege,W.O. and Monahan,J.E.
Genes differentially expressed in human prostate cancer and their use
Patent: WO 0160860-A 22065 23-AUG-2001;
Millennium Predictive Medicine, Inc. (US)
Location/Qualifiers
1..2199
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 61.4%; Score 1579.6; DB 6; Length 2199;
Best Local Similarity 97.1%; Pred. No. 9.1e-247;
Matches 1748; Conservative 0; Mismatches 34; Indels 18; Gaps 13;
QY 695 GHTTTGGAGGAGTACCTCTTTATTTCTGAAGCCCGGCGAGTGTGACCTGTGCGAC 754
Db 187 GCTTGTGCACAGTGTACCTCTCTTTATTTCTGAAGCCCGGCGAGTGTGACCTGTGCGAC 246
QY 755 CAGTCCCAAAACAGATCCAGAAAGAAAGTTTCCAGGTTTGTGCCATCGAAGGACTGGAT 814
Db 247 CAGTCCCAAAACAGATCCAGAAAGAAAGTTTCCAGGTTTGTGCCATCGAAGGACTGGAT 306
QY 815 GCCACGGGTAAACCAACCGGTGACCCAGTCACTGGGAGATTTCACTTAAGGCTGTCTCTTA 874
Db 307 GCCACGGGTAAACCAACCGGTGACCCAGTCACTGGGAGATTTCACTTAAGGCTGTCTCTTA 366
QY 875 AAGTCAACACCTCTTGTGATTTGGCCAGTGGAGGAAGATCTTTGATGATGAACCAACTATC 934
Db 367 AAGTCAACACCTCTTGTGATTTGGCCAGTGGAGGAAGATCTTTGATGATGAACCAACTATC 426
QY 935 ATTAGAAGAGCTTTTACTCTTTGGCAATTATATTGTGGCTTCCGAAATAGCTAAAGAA 994
Db 427 ATTAGAAGAGCTTTTACTCTTTGGCAATTATATTGTGGCTTCCGAAATAGCTAAAGAA 486
QY 995 TCTGCCAAATCTCTGTGATTGTAGACAGGTACTGGCAGCAGCAGCCACCTATGCCATA 1054
Db 487 TCTGCCAAATCTCTGTGATTGTAGACAGGTACTGGCAGCAGCAGCCACCTATGCCATA 546
QY 1055 GCCACTGAGTGAGTGGGGTCTCCAGACCTGCCGCCAGCCCATCACCTGTGTACCG 1114
Db 547 GCCACTGAGTGAGTGGGGTCTCCAGACCTGCCGCCAGCCCATCACCTGTGTACCG 606
QY 1115 TGGCCAGAGGACCTCTCAAACTCAACCTGCTTATCTCTGTGCTCACTGTGAGTCTTGAGGAG 1174
Db 607 TGGCCAGAGGACCTCTCAAACTCAACCTGCTTATCTCTGTGCTCACTGTGAGTCTTGAGGAG 666
QY 1175 AGTTTGCAGAGCTCGAGGCGGGGCATGGHAGAGACCCAGGGAAGAGAGCAACTGTAG 1234
Db 667 AGTTTGCAGAGCTCGAGGCGGGGCATGGHAGAGACCCAGGGAAGAGAGCAACTGTAG 726
QY 1235 GCCACAGTGTGTTTCTGTCAAAAGTAGAAAATGCTCTTACCAGCGATGGAGATCTGTGC 1294
Db 727 GCCACAGTGTGTTTCTGTCAAAAGTAGAAAATGCTCTTACCAGCGATGGAGATCTGTGC 786
QY 1295 TGCCATGTGTTGATGCCAGCCCTCCAGAGAAAAGTCTCTCAGACAGATTTAAGCCCTA 1354
Db 787 TGCCATGTGTTGATGCCAGCCCTCCAGAGAAAAGTCTCTCAGACGCGTATTAAAGCCCTA 846
QY 1355 ATCCAGAAATAGTTTGTAGTGAACCGGTAGTTACTCTGGCCAGGTCGCCAGCTTAACCTAGATT 1414
Db 847 ATCCAGAAATAGTTTGTAGTGAACCGGTAGTTACTCTGGCCAGGTCGCCAGCTTAACCTAGATT 906
QY 1415 AGATGTTGTTGAACATCTACATCCAGATTTGTTATGAGTGTTCCTCAAAATTTCTGTT 1474
Db 907 AGATGTTGTTGAACATCTACATCCAGATTTGTTATGAGTGTTCCTCAAAATTTCTGTT 966
QY 1475 CTACAAGCATGTTGTGCGAGAAAACCTGGAGCAGGCACTCTTAATTTTACTTCAGCCA 1534

QY	1475	CTACAAGCATGTTGTGGCAGAAAACCTGGAGCCAGGCATCTTAATTTTACTTCAGCCA	1534	DEFINITION	Sequence 27924 from Patent WO0160860.		
Db	967	CTACAAGCATGTTGTGGCAGAAAACCTGGAGCCAGGCATCTTAATTTTACTTCAGCCA	1026	ACCESSION	CQ496057		
QY	1535	TGCTACCCCTCTTCTGACTGATGAGCCCGTCATCACAAGGTCCTCTCATCATGTTCCAG	1594	VERSION	CQ496057.1 GI:41461676		
Db	1027	TGCTACCCCTCTTCTGACTGATGAGCCCGTCATCACAAGGTCCTCTCATCATGTTCCAG	1086	KEYWORDS			
QY	1595	TGAGAGCCAGCAGATTGCTTTCTTCTGCGATAGTAAACATTTTCTTGAAACATATGTTT	1654	SOURCE	Homo sapiens (human)		
Db	1087	TGAGAGCCAGCAGATTGCTTTCTTCTGCGATAGTAAACATTTTCTTGAAACATATGTTT	1146	ORGANISM	Homo sapiens		
QY	1655	CACTTAATCACTACCAAAATCTTGAAGACCTGCTTACTCAGACAGCACCAGGTGACA	1714	REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Db	1147	CACTTAATCACTACCAAAATCTTGAAGACCTGCTTACTCAGACAGCACCAGGTGACA	1206	AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
QY	1715	GAAGCAGCAGACAAGATCTTCCAGATCAGCAGGAGACCCGAGGCTCTGCTTCTCCTTA	1774	TITLE	Schlegel, R., Endege, W.O. and Monahan, J.E.		
Db	1207	GAAGCAGCAGACAAGATCTTCCAGATCAGCAGGAGACCCGAGGCTCTGCTTCTCCTTA	1266	GENES	Genes differentially expressed in human prostate cancer and their use		
QY	1775	CACTGCGATGCTGATGAGATCGTGACATGCCCCATTTGGCTTCTCCACATCTGGTTGCA	1834	JOURNAL	Patent: WO 0160860-A 27924 23-AUG-2001;		
Db	1267	CACTGCGATGCTGATGAGATCGTGACATGCCCCATTTGGCTTCTCCACATCTGGTTGCA	1326	FEATURES	Millennium Predictive Medicine, Inc. (US)		
QY	1835	CTCGTCATGATGGGCTCGCTGCATCTCCCTCAGTCCCAAAATTTCTAGTAGCAGAAAGTTC	1894	source	1. .2199		
Db	1327	CTCGTCATGATGGGCTCGCTGCATCTCCCTCAGTCCCAAAATTTCTAG-AGCCAAAGTGTCC	1385	ORIGIN	/organism="Homo sapiens"		
QY	1895	TGCAGAGGCTGTCTATGTCTCTGCTGCTGCCCAAGGACACTCTGTCAGAGGCCATTTTGG	1954	Query Match	61.4%; Score 1579.6; DB 6; Length 2199;		
Db	1386	TGCAGAGGCTGTCTATGTCTCTGCTGCTGCCCAA-GGACACTCCTGCAGAGCCATTTTGG	1444	Best Local Similarity	97.1%; Pred. No. 9.1e-247;		
QY	1955	GTAAGGAAACATTAAGGAAAGCATTGATCTTGTCTGAGGCTCAGAGGCCCTTTTGAT	2014	Matches 1748; Conservative	0; Mismatches 34; Indels 18; Gaps 13;		
Db	1445	GTAAGGAAACATTAAGGAAAGCATTGATCTTGTCTGAGGCTCAGAGGCCCTTTTGAT	1504	QY	695	GTTTGGAGGAGTGTAACCTCTTTTATTCCTGAAGCCCGGGCAGTGTGACCTGTGTCAG	754
QY	2015	AGGCTTCTGATGTCATCTAAGACATTCAGCCCAAGATGCTCCAACTGCGAAATATACC	2074	Db	187	GCTTGTGCACAGTACCTCTTTATTCCTGAAGCCCGGGCAGTGTGACCTGTGTCAG	246
Db	1505	AGGCTTCTGA-GTCAATATAAGACATTCAGCCCAAGATGCTCCAACTGCGAAATATACC	1563	QY	755	CAGTCCCCAAAACAGATCCAGAAAGTTCAGAGTTGTTGCCATCGAAGACTGGAT	814
QY	2075	AACCTTCTCTGAATATATTTGCTATTATTAATTTCTTTTCTTTTCTTAAAGATG	2134	Db	247	CAGTCCCCAAAACAGATCCAGAAAGTTCAGAGTTGTTGCCATCGAAGACTGGAT	306
Db	1564	AA-CTTCTCTGAATATA-TTGTCTTATTATTAATTTCTTTTCTTTTCTTAAAGTA--TG	1619	QY	815	GCCAGGGGTAAACACCGTGACCCAGTCAGTGGCAGATTACATTAAGGCTGTCTCTTTA	874
QY	2135	GCTCTGAATAGATGACATTTTCCATCTGAATGAGATCATATTTAGCCAAATCCAG	2194	Db	307	GCCAGGGGTAAACACCGTGACCCAGTCAGTGGCAGATTACATTAAGGCTGTCTCTTTA	366
Db	1620	GCTCTGAATAGATGACATTTTCCAT-TGAATGGATGCAATTTTCAATTTAGCCAAATCCAG	1678	QY	875	AAGTCAACACCTCTTGCAATGCGCAGTGGAGGAAGATCTTTGATGATGAACCAACTATC	934
QY	2195	TAAATTTTATTAATTAATCTATACATTAATGTTTCTCAGCATAGGACTATGATTCAT	2254	Db	367	AAGTCAACACCTCTTGCAATGCGCAGTGGAGGAAGATCTTTGATGATGAACCAACTATC	426
Db	1679	TAAATTTTATTAATTAATCTATACATTAATGTTTCTCAGCATAGGACTATGATTCAT	1738	QY	935	ATTAGAGAGCTTTTACTCTTTGGGCAATTATATTTGGCTCTCGAATAGCTAAAGAA	994
QY	2255	TAAATTTAAAGTGGAGTCAAAAGCTTAAATGCAATGTTTGTGATTTTTCATTAACAAA	2314	Db	427	ATTAGAGAGCTTTTACTCTTTGGGCAATTATATTTGGCTCTCGAATAGCTAAAGAA	486
Db	1739	TAAATTTAAAGTGGAGTCAAAA-GCTAAATGCAATGTTTGTGATTTTTCATTAACAAA	1794	QY	995	TCTGCAAAATCTCTGTGATTTAGACAGGTAAGTGGCAGCAGCAGCCACCTATGCCATA	1054
QY	2315	ACTTAATTTGCTTGTGTTAAATGATTAAGTCAAGTCAAGTCTTGAGTGGGATTTCTT-GGTAAA	2373	Db	487	TCTGCAAAATCTCTGTGATTTAGACAGGTAAGTGGCAGCAGCAGCCACCTATGCCATA	546
Db	1795	A--TTAAATGCTCTGTAAATAAGTACA--GGGATCTTGAGTGGGATTTCTTGGGTAAA	1850	QY	1055	GCCACTGAGGTGAGTGGGGTCTCCAGCACTGCCCCCAGCCCATACCTGTGTATCCAG	1114
QY	2374	TTATCTTGACATGATGCTCATGATTAATGATTAATGAAATCGCTTTTGACATATCTTTAGAC	2433	Db	547	GCCACTGAGGTGAGTGGGGTCTCCAGCACTGCCCCCAGCCCATACCTGTGTATCCAG	606
Db	1851	TTATCTTGACATGATGCTCTCATGATTAATGATTAATGAAATCGC-TTGACATATCTTTGGAG	1909	QY	1115	TGGCCAGAGGACCTCTCAAACTCAACCTTATCTCTGCTGCTCACTGTGAGTCTCTGAGAG	1174
QY	2434	AGAAAAGTAGCTGAGTGGGGGAAAATTAATAGAGCTTGTGTGATTTAGGGAGTAGCT	2493	Db	607	TGGCCAGAGGACCTCTCAAACTCAACCTTATCTCTGCTGCTCACTGTGAGTCTCTGAGAG	666
Db	1910	AGAAAAGTAGCTGAGTGGGGGAAAATTAATAGAGCTTGTGTGATTTAGGGAGTAGGT	1969	QY	1175	AGTTTSCAGAGGCTGACAGGCGGGGTCATGGAGAAAGACAGGAAAGACGAACTTGAG	1234
LOCUS	CQ496057	2199 bp	DNA	linear	PAT	30-JAN-2004	726

```
QY 1415 AGATCTGTTTGAACATCTATCCACCAATTTGTTATGCACTGTTCCAAATTTCTGTT 1474
Db |||
QY 907 AGATCTGTTTGAACATCTATCCACCAATTTGTTATGCACTGTTCCAAATTTCTGTT 966
Db |||
QY 1475 CTACAGCATGTTGTGTCAGAAACCTGAGACCAAGGATCTTAATTTTACTTCAGCCA 1534
Db |||
QY 967 CTACAGCATGTTGTGTCAGAAACCTGAGACCAAGGATCTTAATTTTACTTCAGCCA 1026
Db |||
QY 1535 TCGTACCCCTCTCTGATGATGACCCGTCATCAAAAGGTCCCTCTCATCATGTTCCAG 1594
Db |||
QY 1027 TCGTACCCCTCTCTGATGATGACCCGTCATCAAAAGGTCCCTCTCATCATGTTCCAG 1086
Db |||
QY 1595 TGAGAGCCAGCATGTTCTTCTCTGTCATAGTAACATTTCTTGGAAATATGTTT 1654
Db |||
QY 1087 TGAGAGCCAGCATGTTCTTCTCTGTCATAGTAACATTTCTTGGAAATATGTTT 1146
Db |||
QY 1655 CACTTAATCACTACCAAAATATCTGGAAGACCTCTTACTCAGACAGCCAGGTGTACA 1714
Db |||
QY 1147 CACTTAATCACTACCAAAATATCTGGAAGACCTCTTACTCAGACAGCCAGGTGTACA 1206
Db |||
QY 1715 GAAGCAGCAGCAAGATCTTCCAGATCAGCAGGAGACCCCGGAGCTCTGCTTCTCCTA 1774
Db |||
QY 1207 GAAGCAGCAGCAAGATCTTCCAGATCAGCAGGAGACCCCGGAGCTCTGCTTCTCCTA 1266
Db |||
QY 1775 CACTGCGATGCTGATGAGATCGTGACATGCCACATTTGGCTTCTCCACATCTGTTGCA 1834
Db |||
QY 1267 CACTGCGATGCTGATGAGATCGTGACATGCCACATTTGGCTTCTCCACATCTGTTGCA 1326
Db |||
QY 1835 CTGCTCATGATGGCTCGCTGTCATCTCCCTCAGTCCCAATTTCTAGTAGCAAGTGTTC 1894
Db |||
QY 1327 CTGCTCATGATGGCTCGCTGTCATCTCCCTCAGTCCCAATTTCTAG - AGCCAAGTGTTC 1385
Db |||
QY 1895 TGCAGAGGCTGTCTATGTGTCCTGGTGCCTCCAGGACATCTCTGAGAGCCATTTTGG 1954
Db |||
QY 1386 TGCAGAGGCTGTCTATGTGTCCTGGTGCCTCCAG - GGACACTCTCTGAGAGCCATTTTGG 1444
Db |||
QY 1955 GTAAGAACACTTACAAGAGGATGTATCTTGTCTGAGGCTCAGAGCCCTTTTGTAT 2014
Db |||
QY 1445 GTAAGAACACTTACAAGAGGATGTATCTTGTGTCAGGCTCAGAGCCCTTTTGTAT 1504
Db |||
QY 2015 AGGCTCTGATGTCAATTAAGACATTCAGGCAAGATGCTCCAACTGCAAAATATACC 2074
Db |||
QY 1505 AGGCTCTGCA - GTCATATATAAGACATTCAGGCAAGATGCTCCAACTGCAAAATATACC 1563
Db |||
QY 2075 AACCTCTCTGAATTAATTTGCTATTAATTAATTTCTTTTCTTTTCTTAAAGATTG 2134
Db |||
QY 1564 AA - CTTCTCTGAAATATA - TTTGCTTATTTATATTTCTTTTCTTTTCTTAAAGTA - TG 1619
Db |||
QY 2135 GCTCTGAATAGATGCACATTTTCCATCTGAAGTGGATGCATATCATTTAGCCCAATCCAG 2194
Db |||
QY 1620 GCTCTGAATAGATGCACATTTTCCAT - TGAACTGGATGCAATTTTATTTAGCCCAATCCAG 1678
Db |||
QY 2195 TAATTTATTTATTAATCTATACATAATATGTTTCTCTCAGCATAGGAGCTATGATTCAT 2254
Db |||
QY 1679 TAATTTATTTATTAATCTATACATAATATGTTTCTCTCAGCATAGGAGCTATGATTCAT 1738
Db |||
QY 2255 TAATTTAAAGTGGAGTCAAAACGCTAAATGCAATGTTTGTGTATTTTCTTAAACACAA 2314
Db |||
QY 1739 TAATTTAAAGTGGAGTCAAAAC - GCTAAATGCAATGT - - - GTGTGTATTTTCTTAAACAA 1794
Db |||
QY 2315 ACTTAATTTGCTTGTAAATAAGTTCAAGTGGATCTTGGAGTGGGATTTCTT - GGTAAA 2373
Db |||
QY 1795 A - - TTAATTTGCTTGTAAATAAGTACA - - GGGATCTTGGAGTGGGATTTCTTGGTAAA 1850
Db |||
QY 2374 TTATCTTGCATCTGAATGTCTCATGATTAATACATATGAAATCGCTTTGACATATCTTTAGAC 2433
Db |||
QY 1851 TTAICTTGCATCTGAATGTCTCTCATTAATACATATGAAATCGC - TTGACATATCTTTGGAC 1909
Db |||
QY 2434 AGAAAAAGTAGCTGAGTGAGGGGAAATATAGAGCTTGTGTACTTTAGGAGTAGCT 2493
Db |||
QY 1910 AGAAAAAGTAGTAGTAGGGGGGAAATATATAGAGGCTGTGTGACTTTAGGGAGTAGGT 1969
Db |||
```

```
RESULT 7
AX814311
LOCUS AX814311 1535 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 4 from Patent WO03064642.
ACCESSION AX814311
VERSION AX814311.1 GI:39103549
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Liou, J.R.
TITLE Regulation of human thymidylate kinase
JOURNAL Patent: WO 03064642-A 4 07-AUG-2003;
Bayer Aktiengesellschaft (DE)
FEATURES
Source
1. .1535
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 56.8%; Score 1459.6; DB 6; Length 1535;
Best Local Similarity 99.7%; Pred. No. 2.8e-227;
Matches 1462; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CGGGGCGAGGGCGGCTCGCTGAGGCGGCCATGGCTTCGCCGCGCGGCTCCTGGCGG 60
Db 70 CGGGGCGAGGGCGGCGTCTGCTGAGGCGGCCATGGCTTCGCCGCGCGGCTCCTGGCGG 129
QY 61 GCCACTGTTCGGGCGCGCTGCTCGGGCGGCGCGGGTCTGCGTGGGCGCCANGGCTCGCC 120
Db 130 GCCACTGTTCGGGCGCGCTGCTCGGGCGGCGCGGGTCTGCGTGGGCGCCATGGCTCGCC 189
QY 121 GTCGCGCTTCTGAGGCTTCCGACTGACACCTGGGCTCACTTCGCCCTAGGCGCGCA 180
Db 190 GCGCGCTTCTGAGGCTTCCGACTGACACCTGGGCTCACTTCGCCCTAGGCGCGCA 249
QY 181 CGGCGCGGCGAGCGAGACGCCCGCGCTCGGGCGGCTGTGGGCGCCCGCGCA 240
Db 250 CGGCGCGGCGAGCGAGACGCCCGCGCTCGGGCGGCTGTGGGCGCCCGCGCA 309
QY 241 GCGCAGCTACTCGCTGTGCTGCGCGCTGACCCCGGACGCCGCGTGCGGGCGCCGGTCCG 300
Db 310 GCGCAGCTACTCGCTGTGCTGCGCGCTGACCCCGGACGCCGCGTGCGGGCGCCGGTCCG 369
QY 301 GGGGCGCGGCTGCAACAGCGCTGCTGACACGCTGCGGCGGCGCCCTTCCAGCGGTG 360
Db 370 GGGGCGCGGCTGCAACAGCGCTGCTGACACGCTGCGGCGGCGCCCTTCCAGCGGTG 429
QY 361 CCAGCTCTCAGGCTGCTCTGTACTGCCGGGCGGCGAGGCCGCGCGCGCGCACAGCAAG 420
Db 430 CCAGCTCTCAGGCTGCTCTGTACTGCCGGGCGGCGAGGCCGCGCGCGCGCACAGCAAG 489
QY 421 CTTCTCTGTCGCGACCCCTTGGATGACCCCTGACACCCCGGCAAGGCTGCTCGAGCTGCT 480
Db 490 CTTCTCTGTCGCGACCCCTTGGATGACCCCTGACACCCCGGCAAGGCTGCTCGAGCTGCT 549
QY 481 GGGGCGCTTCCAGAGGACCAACCGCGCTTGGGCGAGTTCAGGCGCGGCGCGCGCGG 540
Db 550 GGGGCGCTTCCAGAGGACCAACCGCGCTTGGGCGAGTTCAGGCGCGGCGCGCGCGG 609
QY 541 CCAGCTTGGGCGAGCGCTCTTGGAGGTCGAAGAGCGGAGGCGGCTGAGAGTGGGCTGCGC 600
Db 610 CCAGCTTGGGCGAGCGCTCTTGGAGGTCGAAGAGCGGAGGCGGCTGAGAGTGGGCTGCGC 669
QY 601 ACAGTCTGTCGCCGCTCCGAGAGCCCGCTGACACCGGCTGTGCAAGCTTGTGCCAGTTC 660
Db 670 ACAGTCTGTCGCCGCTCCGAGAGCCCGCTGACACCGGCTGTGCAAGCTTGTGCCAGTTC 729
QY 661 GTCGCTTCTCCGAGACCGGAGCGCGCGGCGCTTTTGGAGGAGTGTACTCTCTTTAT 720
Db |||
```

Db 730 CGTGGTCTTCCGGACCGGAAGCGCGCCGGCGGTTTTGGAGGAGTGTAACCTCTTTAT 789
Qy 721 TCCTGAAGCCCGGCGAGTCTTGACCTGGTGCACAGTGCCCAAAACAGATCCAGAAGG 780
Db 790 TCCTGAAGCCCGGCGAGTCTTGACCTGGTGCACAGTGCCCAAAACAGATCCAGAAGG 849
Qy 781 AAAGTTCAGAGTCTTGTCATCGAAGGACTGGATGCCACGGGTAAACACCGGTGACCCA 840
Db 850 AAAGTTCAGAGTCTTGTCATCGAAGGACTGGATGCCACGGGTAAACACCGGTGACCCA 909
Qy 841 GTACGTGGCAGATCACTTAAGGCTGCTCTTTAAAGTCACACCCCTCTTGCAATGGCCA 900
Db 910 GTACGTGGCAGATCACTTAAGGCTGCTCTTTAAAGTCACACCCCTCTTGCAATGGCCA 969
Qy 901 GTGAGGAGATCTTTGATGATGAACCACTATCAATTAGAGAGCTTTTACTCTTTGGG 960
Db 970 GTGAGGAGATCTTTGATGATGAACCACTATCAATTAGAGAGCTTTTACTCTTTGGG 1029
Qy 961 CAATTATATTGTGGCTCCGAAATAGCTAAAGAACTGCCCCAAATCTCTGTGATTGTAGA 1020
Db 1030 CAATTATATTGTGGCTCCGAAATAGCTAAAGAACTGCCCCAAATCTCTGTGATTGTAGA 1089
Qy 1021 CAGTACTGGCAGACGCGCACTATGCAATAGCCACTGAGGTGAGTGGGGTCTCCA 1080
Db 1090 CAGTACTGGCAGACGCGCACTATGCAATAGCCACTGAGGTGAGTGGGGTCTCCA 1149
Qy 1081 GCACCTGCCCCAGCCCATCACCTGTGTACCACTGAGGTGAGTGGGGTCTCCA 1140
Db 1150 GCACCTGCCCCAGCCCATCACCTGTGTACCACTGAGGTGAGTGGGGTCTCCA 1209
Qy 1141 CCTTATCTCTGCTCACTGTGAGTCTCTGAGGAGAGGTGAGAGGCTGAGGCGCGGG 1200
Db 1210 CCTTATCTCTGCTCACTGTGAGTCTCTGAGGAGAGGTGAGAGGCTGAGGCGCGGG 1269
Qy 1201 CATGGAGAAACAGGAGAAAGAGCAAACTTGAGGCCAAACAGTGTGTTTCTCAAAAGGT 1260
Db 1270 CATGGAGAAACAGGAGAAAGAGCAAACTTGAGGCCAAACAGTGTGTTTCTCAAAAGGT 1329
Qy 1261 AGAAATGTCTTACAGCGGATGAGATCTCTGCTGCGATGCTGATGCGAGCGCCCTC 1320
Db 1330 AGAAATGTCTTACAGCGGATGAGATCTCTGCTGCGATGCTGATGCGAGCGCCCTC 1389
Qy 1321 CAGAGAAAAGGTCTCTGAGACAGTATTAAGCCTTAATCCAGATAGTTTTAGTGAACCGTA 1380
Db 1390 CAGAGAAAAGGTCTCTGAGACAGTATTAAGCCTTAATCCAGATAGTTTTAGTGAACCGTA 1449
Qy 1381 GTTACTCTGCGCAGGTGCGACGCTAACTAGATTAGATGTTGTTGAAACATCTACATCC 1440
Db 1450 GTTACTCTGCGCAGGTGCGACGCTAACTAGATTAGATGTTGTTGAAACATCTACATCC 1509
Qy 1441 ACCATTGTTATGTCAGTGTTCCTCAAA 1466
Db 1510 ACCATTGTTATGTCAGTGTTCCTCAAA 1535

RESULT 8
LOCUS AX814308
DEFINITION Sequence 1 from Patent WO03064642.
ACCESSION AX814308
VERSION AX814308.1 GI:39103548
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Liou, J.R.
TITLE Regulation of human thymidylate kinase
JOURNAL Patent: WO 03064642-A 1 07-AUG-2003;
Bayer Aktiengesellschaft (DE)
FEATURES
Location/Qualifiers
source 1..1343

/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 52.0%; Score 1336.6; DB 6; Length 1343;
Best Local Similarity 99.7%; Pred. No. 2.8e-207;
Matches 1339; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 32 ATGSCCTTCGCGCCCGGCTCCTCGCGGGCCACTGTGCGGGCGCTGCTCGGGCGGGC 91
Db 1 ATGSCCTTCGCGCCCGGCTCCTCGCGGGCCACTGTGCGGGCGCTGCTCGGGCGGGC 60
Qy 92 GGGGTCTGCGTGGGGCCATGGCTCCCGCGCTTCGCTCTGAGCTTCCCGACTGC 151
Db 61 GGGGTCTGCGTGGGGCCATGGCTCCCGCGCTTCGCTCTGAGCTTCCCGACTGC 120
Qy 152 ACCCTGGCTCACTTCGCGCTAGGGCCGACGCCCGCGGCGACGAGCGCCCGGACCCC 211
Db 121 ACCCTGGCTCACTTCGCGCTAGGGCCGACGCCCGCGGCGACGAGCGCCCGGACCCC 180
Qy 212 CGCTGCGCGGCTGTTGGGGCGCCCGGAGCGAGCTACTCGCTGTCGTCGCGCTGACC 271
Db 181 CGCTGCGCGGCTGTTGGGGCGCCCGGAGCGAGCTACTCGCTGTCGTCGCGCTGACC 240
Qy 272 CCGGACGCCGCTGCGGGCGCCCGGCTCCGGGCGCGCGCTGACACGCGCTGCTGCAC 331
Db 241 CCGGACGCCGCTGCGGGCGCCCGGCTCCGGGCGCGCGCTGACACGCGCTGCTGCAC 300
Qy 332 CAGCTGCGCGCGCGCGCTTCCAGCGGTGCGAGCTGCTCAGGCTGCTCTGCTACTGCCG 391
Db 301 CAGCTGCGCGCGCGCGCTTCCAGCGGTGCGAGCTGCTCAGGCTGCTCTGCTACTGCCG 360
Qy 392 GGGCGCGCGCGCGCGCACAGCAAGGCTTCTGCTGCGCGCACCCCTCGATGACCT 451
Db 361 GGGCGCGCGCGCGCGCACAGCAAGGCTTCTGCTGCGCGCACCCCTCGATGACCT 420
Qy 452 GACACCGCGCAAGCGTGTCTGAGCTGCTGGGCGCTGCGAGGAGCACACCGCGCAC 511
Db 421 GACACCGCGCAAGCGTGTCTGAGCTGCTGGGCGCTGCGAGGAGCACACCGCGCAC 480
Qy 512 TTGGGCGAGTTCAGAGCGCGCACCGCGCGCGCGCTGTTGGAGGCTGTTGGAGGTCAA 571
Db 481 TTGGGCGAGTTCAGAGCGCGCACCGCGCGCGCGCTGTTGGAGGCTGTTGGAGGTCAA 540
Qy 572 GAGCGAGGCGGTGCGAGGTGGGCTGCGCACAGGTGTTGCCCGCTCCGAGAGCCCGCTG 631
Db 541 GAGCGAGGCGGTGCGAGGTGGGCTGCGCACAGGTGTTGCCCGCTCCGAGAGCCCGCTG 600
Qy 632 CAGCGGTGCTGCGAGCTTGGCCAGTTCGCTGGTCTTCCCGAGCCGGGAAGCGCCCG 691
Db 601 CAGCGGTGCTGCGAGCTTGGCCAGTTCGCTGGTCTTCCCGAGCCGGGAAGCGCCCG 660
Qy 692 GCGTTTTGAGGAGTGTACTCTTTATTCTGAGCGCGCGGCGAGTGTGACCTGGTC 751
Db 661 GCGTTTTGAGGAGTGTACTCTTTATTCTGAGCGCGCGGCGAGTGTGACCTGGTC 720
Qy 752 GACAGTGCGCCAAAACAGATCCAGAAAGGAAAGTTCAGGTTGTTGCCATCGAAGGACTG 811
Db 721 GACAGTGCGCCAAAACAGATCCAGAAAGGAAAGTTCAGGTTGTTGCCATCGAAGGACTG 780
Qy 812 GATGCCACGGGTAAACCAACGCTGACCCAGTCCAGTGGCAGATTCATTTAAGGCTGTCCTC 871
Db 781 GATGCCACGGGTAAACCAACGCTGACCCAGTCCAGTGGCAGATTCATTTAAGGCTGTCCTC 840
Qy 872 TTAAAGTACACCCCTCTTTCGATTCGCGAGTGGAGAGATCTTTGATGATGAACCACT 931
Db 841 TTAAAGTACACCCCTCTTTCGATTCGCGAGTGGAGAGATCTTTGATGATGAACCACT 900
Qy 932 ATCAATTAGAAGAGCTTTTACTCTTTGGGCAATTATATTGTCGCTCCGAAATAGCTAA 991
Db 901 ATCAATTAGAAGAGCTTTTACTCTTTGGGCAATTATATTGTCGCTCCGAAATAGCTAA 960

QY	992	GAATCTGCCAAATCTCTCTGATGTAGACAGTACTGGCAGACAGCGCCACCTATGCC	1051
Db	961	GAATCTGCCAAATCTCTCTGATGTAGACAGTACTGGCAGACAGCGCCACCTATGCC	1020
QY	1052	ATAGCCACTGAGTGTAGTGGGGTCTCCAGACCTCTCCCGCCAGCCCATCACCTGTGTAC	1111
Db	1021	ATAGCCACTGAGTGTAGTGGGGTCTCCAGACCTCTCCCGCCAGCCCATCACCTGTGTAC	1080
QY	1112	CAGTGGCCAGACCTGCTCAACCTGACCTTATCTGCTGCTCACTGTGAGTCTCTGAG	1171
Db	1081	CAGTGGCCAGACCTGCTCAACCTGACCTTATCTGCTGCTCACTGTGAGTCTCTGAG	1140
QY	1172	GAGAGGTTCGACAGGCTGACAGGCGCGGGCATCGAGAACCCAGGGAAGCAGAACTT	1231
Db	1141	GAGAGGTTCGACAGGCTGACAGGCGCGGGCATCGAGAACCCAGGGAAGCAGAACTT	1200
QY	1232	GAGGCAACAGTGTGTTTCTGTAAGAGTGAATGCTTACACGCGGATGGAGATCTCT	1291
Db	1201	GAGGCAACAGTGTGTTTCTGTAAGAGTGAATGCTTACACGCGGATGGAGATCTCT	1260
QY	1292	GGCTGCCATGTTGATGCGAGCGCCCTCCAGAGAAAGGTCTCTGCAGACAGTATTAGC	1351
Db	1261	GGCTGCCATGTTGATGCGAGCGCCCTCCAGAGAAAGGTCTCTGCAGACAGTATTAGC	1320
QY	1352	CTAATCCAGAAATAGTTTTAGTGA	1374
Db	1321	CTAATCCAGAAATAGTTTTAGTGA	1343

RESULT 9

BC016969

LOCUS

BC016969 Homo sapiens, clone IMAGE:4428577, mRNA, partial cds.

ACCESSION

BC016969

VERSION

BC016969.1 GI:16877443

KEYWORDS

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1720)

Strausberg,R.

Direct Submission

Submitted (05-NOV-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs@mail.nih.gov

Tissue Procurement: DCTD/DP

CDNA Library Prepared by: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mcddpaxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 27 Row: n Column: 24.

Location/Qualifiers

1. .1720

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4428577"

/tissue_type="Prostate, adenocarcinoma."

/clone_lib="NIH MGC 91"

/lab_host="DH10E"

/note="Vector: pCMV-SPORT6"

QY	1205	GAGAAGCAGGGAAGAGCAGAACTTGAGGCGCAACAGTGTGTTTCGTCAAAGGTAGAA	1264
Db	1	GAGAAGCAGGGAAGAGCAGAACTTGAGGCGCAACAGTGTGTTTCGTCAAAGGTAGAA	60
QY	1265	ATGTCCTTACGAGCGGATGGAGAACTCTGGCTGCCATGTGGTTCATGCCGCCCTCCAGA	1324
Db	61	ATGTCCTTACGAGCGGATGGAGAACTCTGGCTGCCATGTGGTTCATGCCGCCCTCCAGA	120
QY	1325	GAAAAGTCTCTGACAGACAGTATTAAAGCCTAATCCAGAATAAGTTTAACTGAACCGTAGTAA	1384
Db	121	GAAAAGTCTCTGACAGACAGTATTAAAGCCTAATCCAGAATAAGTTTAACTGAACCGTAGTAA	180
QY	1385	CTCTGGCCAGGTGCCAGTCTAACTAGATTAGATGTTTGTGAAACATCTACATCCACCA	1444
Db	181	CTCTGGCCAGGTGCCAGTCTAACTAGATTAGATGTTTGTGAAACATCTACATCCACCA	240
QY	1445	TTTGTTATGAGTGTTCCTCCAAATTTCTGTTCTACAGCATGTTGTGGCGAANAATCGG	1504
Db	241	TTTGTTATGAGTGTTCCTCCAAATTTCTGTTCTACAGCATGTTGTGGCGAANAATCGG	300
QY	1505	AGACCGCATCTTAATTTTACTTCAGCATCTGACCTCTCTGACTGATGAGCCCGTC	1564
Db	301	AGACCGCATCTTAATTTTACTTCAGCATCTGACCTCTCTGACTGATGAGCCCGTC	360
QY	1565	ATCACAAGGTCCCTCTCATCATGTTCCAGTGAGAGGCGCAGATGCTTCTTCTCTGGC	1624
Db	361	ATCACAAGGTCCCTCTCATCATGTTCCAGTGAGAGGCGCAGATGCTTCTTCTCTGGC	420
QY	1625	ATAGTAAACATTTTCTTGGAAACATATGTTTCACTTAATCACTACCAATATCTGGAAC	1684
Db	421	ATAGTAAACATTTTCTTGGAAACATATGTTTCACTTAATCACTACCAATATCTGGAAC	480
QY	1685	CTGTCTTACTCAGACAGCAGGATGACAGAGCAGCAGACAGATCTTCAGATCAGC	1744
Db	481	CTGTCTTACTCAGACAGCAGGATGACAGAGCAGCAGACAGATCTTCAGATCAGC	540
QY	1745	AGGAGACCCCGGAGCCT	1804
Db	541	AGGAGACCCCGGAGCCT	600
QY	1805	CCACATTCGCTCTTCCACATCTGTTGACATCTGATGAGGCGCTGCTGATCTCTCTCT	1864
Db	601	CCACATTCGCTCTTCCACATCTGTTGACATCTGATGAGGCGCTGCTGATCTCTCTCT	660
QY	1865	CAGTCCCAATTTCTAGTAGCAGCAAGTTCTCTCAGAGGCTGCTATGTTCTCTGGTGGC	1924
Db	661	CAGTCCCAATTTCTAGTAGCAGCAAGTTCTCTCAGAGGCTGCTATGTTCTCTGGTGGC	719
QY	1925	CAAGGACACCTCTCTCAGAGCCTATTTTGGTGAAGAAACATTAAGAGGCAATTC	1984
Db	720	CAA-GGACACCTCTCTCAGAGCCTATTTTGGTGAAGAAACATTAAGAGGCAATTC	778
QY	1985	CTTGTGCTGAGGCTCAGAGCCCTTTTGTAGTGGCTTCTGATGCTCAATTAAGAGCAATTC	2044
Db	779	CTTGTGCTGAGGCTCAGAGCCCTTTTGTAGTGGCTTCTGATGCTCAATTAAGAGCAATTC	837
QY	2045	AAAGCCAGATGCTCCAACTGCAAAATATACCAACCTTCTCTGAATTAATTTTCTTATTT	2104
Db	838	AAAGCCAGATGCTCCAACTGCAAAATATATACCAACCTTCTCTGAATTAATTTTCTTATTT	897

CDS

Query Match
Best Local Similarity 98.9%;
Matches 1275; Conservative 0; Mismatches 7; Indels 7; Gaps 7;

ORIGIN

cl. .177
/codon_start=1
/product="unknown (protein for IMAGE:4428577)"
/protein_id="AAH16969.1"
/db_xref="GI:16877444"
/translation="EKTRBEAELEANSVFRQKVBMSYORMENPGCHVVDASPSREKVL
QTVLSLIQNPSFEP"

misc_feature	28538..28962	/note="similar to EST AI955710 (NID:95748020) wt37e05.xl"
misc_feature	28541..28968	/note="similar to EST AI955710 (NID:95748020) wt37e05.xl"
repeat_region	29077..29381	/rpt_family="L1"
repeat_region	29700..29817	/rpt_family="Alu"
repeat_region	29819..29989	/rpt_family="GA-rich"
repeat_region	30022..30479	/rpt_family="L1"
Query Match	44.4%;	Score 1141.4; DB 9; Length 118808;
Best Local Similarity	98.2%;	Pred. No. 1.9e-175;
Matches 1228;	Conservative 0;	Mismatches 16; Indels 7; Gaps 7;
QY	1243	TGTGTTTCGTAAGAGGTAGAAATGTCCTACAGCGGATGGAGATCCCTGGCTGCCATGT 1302
DB	49421	TTTGCTGTCCCGAGGTAGAAATGTCCTACAGCGGATGGAGATCCCTGGCTGCCATGT 49362
QY	1303	GGTTGATGCCAGCCCTCCAGAGAAAGGTCTCGAGAGAGGTCTCGAGAGAGGTATTAAGCTATATCCAGAA 1362
DB	49361	GGTTGATGCCAGCCCTCCAGAGAAAGGTCTCGAGAGAGGTCTCGAGAGAGGTATTAAGCTATATCCAGAA 49302
QY	1363	TAGTTTTAGTGAACCGTAGTTACTCTGCGCAGGTGCGAGTCTAACTAGATTAGATGTTG 1422
DB	49301	TAGTTTTAGTGAACCGTAGTTACTCTGCGCAGGTGCGAGTCTAACTAGATTAGATGTTG 49242
QY	1423	TTTGAACAATCTACATCCACCATTTGTTATGAGTGTCTCCAAATTTCTGTTCTACAAGC 1482
DB	49241	TTTGAACAATCTACATCCACCATTTGTTATGAGTGTCTCCAAATTTCTGTTCTACAAGC 49182
QY	1483	ATGTTCTGTGGCAAAAACCTGGAGACCGAGCATTTAAATTTTACTTTCAGCCATCGTACCC 1542
DB	49181	ATGTTCTGTGGCAAAAACCTGGAGACCGAGCATTTAAATTTTACTTTCAGCCATCGTACCC 49122
QY	1543	TCCTTCAGTGTGGACCGGTATCAACAAGGTCCCTCTCATCATGTTCCAGTGAGAGGC 1602
DB	49121	TCCTTCAGTGTGGACCGGTATCAACAAGGTCCCTCTCATCATGTTCCAGTGAGAGGC 49062
QY	1603	CAGCGATTGCTTTCTTCTGGGATAGTAACATTTCTTGGACATATGTTTCACTTAAT 1662
DB	49061	CAGCGATTGCTTTCTTCTGGGATAGTAACATTTCTTGGACATATGTTTCACTTAAT 49002
QY	1663	CACCTACCAATATCTGGAAGACCTGCTTACTCAGACAGCACCAGGTGTACAGAAGCAGC 1722
DB	49001	CACCTACCAATATCTGGAAGACCTGCTTACTCAGACAGCACCAGGTGTACAGAAGCAGC 48942
QY	1723	AGACAAGATTTCCAGATCAGCAGGAGACCCCGGAGCCTCTGCTTCTCTACACTGGCA 1782
DB	48941	AGACAAGATTTCCAGATCAGCAGGAGACCCCGGAGCCTCTGCTTCTCTACACTGGCA 48882
QY	1783	TGCTGATGAGATCGTGACATGCCACATTTGGCTTCTTCCACATCTGTTGTCACCTCGTCAT 1842
DB	48881	TGCTGATGAGATCGTGACATGCCACATTTGGCTTCTTCCACATCTGTTGTCACCTCGTCAT 48822
QY	1843	GATGGGCTCGTGCATCTCCCTCAGTCCCAAAATTTAGTAGCCCAAGTGTTCCTGCAGAG 1902
DB	48821	GATGGGCTCGTGCATCTCCCTCAGTCCCAAAATTTAGTAGCCCAAGTGTTCCTGCAGAG 48763
QY	1903	CTGCTATGTTCTTGCTGCCCAAGGACACTCTCTGCAGAGCCATTTTGGGTAGAGAA 1962
DB	48762	CTGCTATGTTCTTGCTGCCCAAGGACACTCTCTGCAGAGCCATTTTGGGTAGAGAA 48704
QY	1963	CACCTACCAAGAGGATTCATCTGTGTCTGAGGCTCAGAGCCCTTTTGTATAGGCTTCT 2022
DB	48703	CACCTACCAAGAGGATTCATCTGTGTCTGAGGCTCAGAGCCCTTTTGTATAGGCTTCT 48644
QY	2023	GATGTCATTCATAAAGACATTCAGCCCAAGATGTCCTCAACTGCATAATATATCAACCTTCT 2082
DB	48643	GA-GTCATATATAAGACATTCAGCCCAAGATGTCCTCAACTGCATAATATATCAACCTTCT 48585

SOURCE	Unclassified.	ORGANISM	Unknown.
	1 (bases 1 to 20566)		Unknown.
REFERENCE	Wei,M.-H., Kechum,K.A., Beasley,E.M. and DiFrancesco,V.	AUTHORS	Isolated human kinase proteins
	Patent: US 664087-A 3 16-DEC-2003;		JOURNAL
FEATURES	Location/Qualifiers	source	1..20966
	/organism="unknown"		/mol_type="genomic DNA"
ORIGIN	Query Match	44.0%;	Score 1130.4; DB 6; Length 20966;
	Best Local Similarity		98.1%; Pred. No. 1.1e-173;
Matches	1228; Conservative	0;	Mismatches 16; Indels 8; Gaps 8;
	1243		TGTGTTTCGTCAAAAGGTAGAAATGTCCTACACGCGGATGGAGAATCCTGGCTGCCATGT 1302
QY	17724	DB	TTTGCCTGTCCCAGGGTAGAAATGTCTACACGCGGATGGAGAATCCTGGCTGCCATGT 17783
	1303	QY	GTTGATGCCAGCCCTCCAGAGAAAAGGTCCCTGCAGACAGATATTAGCCCTAATCCAGAA 1362
QY	17784	DB	GGTGTATGCCAGCCCTCCAGAGAAAAGGTCTCTGCAGACGTAATTAGCCCTAATCCAGAA 17843
	1363	QY	TAGTTTTAGTGAACCGTAGTTACTCTGGCCAGGTGCCAAGTCTAACTAGATAGTTG 1422
QY	17844	DB	TAGTTTTAGTGAACCGTAGTTACTCTGGCCAGGTGCCAAGTCTAACTAGATAGTTG 17903
	1423	QY	TTTGAACATCTACATCCACATTTGTTATGTCAGTGTTCCTCAAAATTTCTGTCTACAAAGC 1482
QY	17904	DB	TTTGAACATCTACATCCACATTTGTTATGTCAGTGTTCCTCAAAATTTCTGTCTACAAAGC 17963
	1483	QY	ATGTTGTGTGCAGAAAACCTGGACACGAGCATCTTAATTTTACTCTCAGCCATCGTACCC 1542
QY	17964	DB	ATGTTGTGTGCAGAAAACCTGGACACGAGCATCTTAATTTTACTCTCAGCCATCGTACCC 18023
	1543	QY	TC-TTCTGACTGATGGACCCGTCTATCAAAAGTCCCTCTCATCATGTTCCAGTGAGAGG 1601
QY	18024	DB	TC-TTCTGACTGATGGACCCGTCTATCAAAAGTCCCTCTCATCATGTTCCAGTGAGAGG 18083
	1602	QY	CCAGCATGCTTTCTTCTGGCATAGTAACATTTCTTGGACACATATGTTTCACTTAA 1661
QY	18084	DB	CCAGCATGCTTTCTTCTGGCATAGTAACATTTCTTGGACACATATGTTTCACTTAA 18143
	1662	QY	TCATACCAATATCTGGAACCTGTCTTACTCAGACAGCACAGGTGTACAGAAGCAG 1721
QY	18144	DB	TCATACCAATATCTGGAACCTGTCTTACTCAGACAGCACAGGTGTACAGAAGCAG 18203
	1722	QY	CAGACAAGATCTTCAGATCAGCGGAGACCCCGGAGCCTCTGCTTCTCCTACACTGGC 1781
QY	18204	DB	CAGACAAGATCTTCAGATCAGCGGAGACCCCGGAGCCTCTGCTTCTCCTACACTGGC 18263
	1782	QY	ATGCTGATGAGATCGTGACATGCCACATTTGCTTCTTCCACATCTGGTTGCACTCGCA 1841
QY	18264	DB	ATGCTGATGAGATCGTGACATGCCACATTTGCTTCTTCCACATCTGGTTGCACTCGCA 18323
	1842	QY	TGATGGGCTCGCTGATCTCCCTCAGTCCCAAAATCTAGTAGCCAAAGTGTCTCTCCAGAG 1901
QY	18324	DB	TGATGGGCTCGCTGATCTCCCTCAGTCCCAAAATCTAGTAGCCAAAGTGTCTCTCCAGAG 18382
	1902	QY	GCTGTCTATGTCCTGGTGCCTCAAGGACACTCTCTGAGAGCCATTTTGGGTAAAGGA 1961
QY	18383	DB	GCTGTCTATGTCCTGGTGCCTCAAGGACACTCTCTGAGAGCCATTTTGGGTAAAGGA 18441
	1962	QY	ACACTTACAAAGAGCATTTGATCTTGTGTCTGAGGCTCAGAGCCCTTTTGTATAGCTTTC 2021
QY	18442	DB	ACACTTACAAAGAGCATTTGATCTTGTGTCTGAGGCTCAGAGCCCTTTTGTATAGCTTTC 18501
	2022	QY	TGATGTCAATTCATAAAGACATTTCAAGCCCAAGATGCTCCAACTGTGCAAAATATACCACTTTC 2081
QY	18502	DB	TGA-GTCAATATATAAAGACATTTCAAGCCCAAGATGCTCCAACTGTGCAAAATATACCACTTTC 18560

QY 272 CCGGACCGGCTCGGGGCGCGGTTCGGGCGCGGCTGACACAGCGCTCTGCTGAC 331
Db 235 ----- 234
QY 332 CAGCTGGCGCGCGGCTTCACAGGCTGTCAGCTGCTCAGGCTGCTCTGCTACTGCGCG 391
Db 235 ----- 234
QY 392 GCGGCGCAGGCGCGCGGCGACACAAAGGCTTCCTGTGCGCGACCCCTCTGGATGACCT 451
Db 235 -----GCGGCGCGCGCACACAAAGGCTTCCTGTGCGCGACCCCTCTGGATGACCT 285
QY 452 GACACCGCGCAAGCGCTCTGAGCTGCTGGCGCCCTGCGAGGAGGACACGCGCCGAC 511
Db 286 GACACCGCGCAAGCGCTGCTGAGCTGCTGGCGCCCTGCGAGGAGGACACGCGCCGAC 345
QY 512 TTGGCGGAGTTCCGAGGCGCGACCGCGCGGCGAGCTGTCGACAGCGCTCTGGAGGTCAA 571
Db 346 TTGGCGGAGTTCCGAGGCGCGACCGCGCGGCGAGCTGTCGAGCGCTCTGGAGGTCAA 405
QY 572 GACGCGAGCGGCTGCAAGTGGGTGCGCACAGTGTGTCGCGGAGCGCCCGCTG 631
Db 406 GACGCGAGCGGCTGCAAGTGGGTGCGCACAGTGTGTCGCGGAGCGCCCGCTG 465
QY 632 CACCGGTGTGTCAGACTTGGCCAGTTCGCTGCTTCGCGAGCGGAGCGCGCGG 691
Db 466 CACCGGTGTGTCAGACTTGGCCAGTTCGCTGCTTCGCGAGCGGAGCGCGCGG 525
QY 692 GCGGTTTTGGAGGAGTGTACTCTCTTATTCCTGAAGCGCGGCGAGTGTGACTGGTC 751
Db 526 GCGGTTTTGGAGGAGTGTACTCTCTTATTCCTGAAGCGCGGCGAGTGTGACTGGTC 585
QY 752 GACAGTGCCCAAAACAGATCCAGAAAGAAAGTTCAGGTTGTTGCCATCGAAGACTG 811
Db 586 GACAGTGCCCAAAACAGATCCAGAAAGAAAGTTCAGGTTGTTGCCATCGAAGACTG 645
QY 812 GATGCCACGGGTAAACACCGGTGACCGAGTTCAGTGCAGATTCACCTAAGCTGTCTC 871
Db 646 GATGCCACGGGTAAACACCGGTGACCGAGTTCAGTGCAGATTCACCTAAGCTGTCTC 705
QY 872 TTAAGTCAACACCTCTTGATGCGGAGTGGAGGAAGTCTTTGATGATGAACCAACT 931
Db 706 TTAAGTCAACACCTCTTGATGCGGAGTGGAGGAAGTCTTTGATGATGAACCAACT 765
QY 932 ATCAITAGAGAGCTTTTACTCTTTGGGCAATTATTTGGCTCTCCGAAATAGCTAAA 991
Db 766 ATCAITAGAGAGCTTTTACTCTTTGGGCAATTATTTGGCTCTCCGAAATAGCTAAA 825
QY 992 GAATCTGCCAAATCTCTGTGATTGTAGACAGTACTGCGACAGCGCGGACCTATGCC 1051
Db 826 GAATCTGCCAAATCTCTGTGATTGTAGACAGTACTGCGACAGCGCGGACCTATGCC 885
QY 1052 ATAGCCACTGAGTGAGTGGGGTCTCCAGACCTGCGCCCGCCCATCACCTGTGTAC 1111
Db 886 ATAGCCACTGAGTGAGTGGGGTCTCCAGACCTGCGCCCGCCCATCACCTGTGTAC 945
QY 1112 CAGTGCCAGAGAGCTGTCAAACCTGACTTATCTGTGTCTCTGTGAGTCTGTAG 1171
Db 946 CAGTGCCAGAGAGCTGTCAAACCTGACTTATCTGTGTCTCTGTGAGTCTGTAG 1005
QY 1172 GAGAGGTTGAGAGGTTGAGGCGCGGCGCATGAGAGGAGGAGGAGGAGGAGGAGT 1231
Db 1006 GAGAGGTTGAGAGGTTGAGGCGCGGCGCATGAGAGGAGGAGGAGGAGGAGGAGT 1065
QY 1232 GAGGCCAACAGTGTGTTTCTCAAAAGGTAGAAATGTCTTACAGCGGATGGAATCCT 1291
Db 1066 GAGGCCAACAGTGTGTTTCTCAAAAGGTAGAAATGTCTTACAGCGGATGGAATCCT 1125
QY 1292 GGTGTCATGTGTTGATGCGACCGCTCCAGAGAAAGGTCTGAGACAGTATTAAAG 1351
Db 1126 GGTGTCATGTGTTGATGCGACCGCTCCAGAGAAAGGTCTGAGACAGTATTAAAG 1185

QY 1352 CTATCCAGAATAGTTTTAGTGAACCGTAG 1381
Db 1186 CTATCCAGAATAGTTTTAGTGAACCGTAG 1215
RESULT 14
AK127983
LOCUS
DEFINITION
1609 bp mRNA linear PRI 19-FEB-2004
Homo sapiens cDNA FLJ46098 fis, clone TESTI2021654, weakly similar
to Mus musculus thymidylate kinase family LPS-inducible member
(Tyk1).
ACCESSION
AK127983
VERSION
AK127983.1 GI:34535131
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
AUTHORS
Kanehori,K., Ishibashi,T., Chiba,Y., Fujimori,K., Hiraoka,S.,
Tanai,H., Watanabe,S., Ishida,S., Ono,Y., Hotuta,T., Watanabe,M.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamanoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Takahashi-Fujii,A., Oshima,A., Sugiyama,A.,
Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
and Isogai,T.
TITLE
NEDO human cDNA sequencing project
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1609)
AUTHORS
Isogai,T. and Yamamoto,J.
TITLE
Direct Submission
JOURNAL
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
[E-mail:genomics@hri.co.jp, Tel.81-438-52-3975, Fax:81-438-52-3986]
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: Reverse Proteomics Research Institute, HRI and
RAB.
FEATURES
source
1. .1609
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TESTI2021654"
/tissue_type="testis"
/clone_lib="TESTI2"
/note="cloning vector: pME18SFL3"
212. 1123
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC87217.1"
/db_xref="GI:34535132"
/translation="MAFARLLRGLSPGLGRGVGACGAMAPRRFVLELPDCTLAH
FALGDADGADADAPDLRALALGPPERSYSLCVPTPDAGCGARVRAHLRHLHLQ
RGPQRCOLLRLLCTCGGAGGAGQGLRDLDDPDPTQALLLGLGACGAEPRPH
LGEFADPRGLQWLWEVDQRQLVGCAQVVPVPEPLFVPLDSSVFPDREH
ARAVLEECTSFIPAEARVLDLVDCPKQTKGKFQVAIEGLDATGKTVTQSVADSL
KAVPQITLCTSGQRTCSNLTLSCSL"
CDS
212. 1123
/note="unnamed protein product"
ORIGIN
Query Match 34.6%; Score 889.2; DB 9; Length 1609;
Best Local Similarity 83.4%; Pred. No. 1.5e-134;
Matches 1121; Conservative 0; Mismatches 3; Indels 220; Gaps 1;
QY 1 CGGGGCGGAGGCGGCGTCTGAGCGGCCCATGCGCTCGCCCGCGCTCTGCGCGG 60
Db 181 CGGGGCGGAGGCGGCGTCTGAGCGGCCCATGCGCTCTGCGCGCGGCTCTGCGCGG 240

Qy	61	GCAC	TGT	CGGG	CCG	CT	CT	CT	CGGG	GGG	CGGG	GT	CT	GG	CT	GGG	GC	AT	TGG	CT	CG	CC	120
Db	241	GCAC	TGT	CGGG	CCG	CT	CT	CT	CGGG	GGG	CGGG	GT	CT	GG	CT	GGG	GC	AT	TGG	CT	CG	CC	300
Qy	121	GT	CG	CG	CT	T	CG	AG	CT	TC	CG	AG	CT	GC	AC	CT	GC	T	CAC	TT	CG	CC	180
Db	301	GC	CG	CG	CT	T	CG	AG	CT	TC	CG	AG	CT	GC	AC	CT	GC	T	CAC	TT	CG	CC	360
Qy	181	CG	CC	CC	CG	CG	AG	CG	CC	CC	CG	AG	CG	CC	CC	CG	AG	CG	CT	T	T	GG	240
Db	361	CG	CC	CC	CG	CG	AG	CG	CC	CC	CG	AG	CG	CC	CC	CG	AG	CG	CT	T	T	GG	420
Qy	241	CG	CG	AG	CT	ACT	CG	CT	GT	CG	CG	AG	CG	CC	CG	AG	CG	CC	CG	AG	CG	CC	300
Db	421	CG	CG	AG	CT	ACT	CG	CT	GT	CG	CG	AG	CG	CC	CG	AG	CG	CC	CG	AG	CG	CC	480
Qy	301	GG	CG	CG	CG	CT	GC	AC	CAG	CG	CG	CT	GC	AC	CAG	CG	CG	CG	CT	GC	AC	CAG	360
Db	481	GG	CG	CG	CG	CT	GC	AC	CAG	CG	CG	CT	GC	AC	CAG	CG	CG	CG	CT	GC	AC	CAG	540
Qy	361	CC	AG	CT	GC	T	CG	AG	CT	GC	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	420
Db	541	CC	AG	CT	GC	T	CG	AG	CT	GC	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	CG	600
Qy	421	CT	CT	CT	CT	CG	CG	AG	CG	CC	CG	AG	CG	CC	CG	AG	CG	CC	CG	AG	CG	CC	480
Db	601	CT	CT	CT	CT	CG	CG	AG	CG	CC	CG	AG	CG	CC	CG	AG	CG	CC	CG	AG	CG	CC	660
Qy	481	GG	CG	CG	CT	GC	AC	CAG	CG	CG	CG	CT	GC	AC	CAG	CG	CG	CG	CT	GC	AC	CAG	540
Db	661	GG	CG	CG	CT	GC	AC	CAG	CG	CG	CG	CT	GC	AC	CAG	CG	CG	CG	CT	GC	AC	CAG	720
Qy	541	CC	AG	CT	GC	AG	CG	CG	CT	CG	AG	CT	GC	AG	CG	CG	CG	CT	CG	AG	CT	GC	600
Db	721	CC	AG	CT	GC	AG	CG	CG	CT	CG	AG	CT	GC	AG	CG	CG	CG	CT	CG	AG	CT	GC	780
Qy	601	AC	AG	CT	GC	T	CG	CG	AG	CG	CC	CG	CT	GC	AC	CC	CG	CT	GC	AC	CC	CG	660
Db	781	AC	AG	CT	GC	T	CG	CG	AG	CG	CC	CG	CT	GC	AC	CC	CG	CT	GC	AC	CC	CG	840
Qy	661	CG	TG	CT	TC	CG	AG	CG	CG	CG	CG	CG	CT	TC	CG	AG	CG	CG	CT	TC	CG	AG	720
Db	841	CG	TG	CT	TC	CG	AG	CG	CG	CG	CG	CG	CT	TC	CG	AG	CG	CG	CT	TC	CG	AG	900
Qy	721	TC	CT	CA	AG	CC	CG	CG	AG	CT	GC	AG	CT	GC	AG	CC	AA	AG	CA	GC	AT	CG	780
Db	901	TC	CT	CA	AG	CC	CG	CG	AG	CT	GC	AG	CT	GC	AG	CC	AA	AG	CA	GC	AT	CG	960
Qy	781	AA	AG	TC	CA	AG	TC	CA	AG	TC	CA	AG	TC	CA	AG	TC	CA	AG	TC	CA	AG	TC	840
Db	961	AA	AG	TC	CA	AG	TC	CA	AG	TC	CA	AG	TC	CA	AG	TC	CA	AG	TC	CA	AG	TC	1020
Qy	841	GT	CA	GT	GC	AG	TC	CA	GT	GC	AG	TC	CA	GT	GC	AG	TC	CA	GT				

	Qy	Dy	Qy	Dy	Qy	Dy	Qy	Dy	Qy	Dy	RESULT 15	BC057565	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK	COMMENT													
	1141	CCTTATCTCTGCTCCTACTGAGTCCTGAGTGAGGTTCGAGGAGAGAGTTCCGAGGCGTGCAGGCCGGGG	12020	1101	CCTTATCTCTGCTCCTACTGAGTCCTGAGTGAGGTTCGAGGAGAGAGTTCCGAGGCGTGCAGGCCGGGG	11650	1201	CATGGAGAAGACACGAGGAAGACAGAACTTGAGGCCCAACAGTGTGTTTCGTCAAAGAGT	12650	1161	CATGGAGAAGACACGAGGAAGACAGAACTTGAGGCCCAACAGTGTGTTTCGTCAAAGAGT	12250	1261	AGAAATGCTCTACCAGCGGATGGAGAACTCTGGCTGCGCATGTGGTTGATGCCAGCCCCTC	13220	1221	AGAAATGCTCTACCAGCGGATGGAGAACTCTGGCTGCGCATGTGGTTGATGCCAGCCCCTC	12800	1321	CAGAGAAAAGGTCTCTGCAGACAGT	1344	1381	CAGAGAAAAGGTCTCTGCAGACAGT	1304	BC057565	Mus musculus thymidylate kinase family LPS-inducible member, mRNA (CDNA clone MGC:16948 IMAGE:6401867), complete cds.	BC057565	1	GI:34785824	MGC.	Mus musculus (house mouse)	Mus musculus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	1 (bases 1 to 3028)	Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L.H., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, B., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.W., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Udwin, T.B., Toshikiyuki, S., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.N., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettunen, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywicki, M.I., Skalska, U., Smalls, D.E., Schercher, A., Schein, J.E., Jones, S.J. and Marra, M.A.	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	12477932	2 (bases 1 to 3028)	Straussberg, R.	Direct Submission	Submitted (03-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgapsh@mail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa CDNA Library Preparation: M. Bento Soares, University of Iowa CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94304 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@axil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

clone distribution. MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 125 Row: b Column: 17
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 15805017.

FEATURES

source

1..3028
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="MGC:66848 IMAGE:6401867"
 /tissue_type="Mouse, brain 12.5 dpc"
 /clone_lib="NTH BMAP_FCO"
 /lab_host="DH10B"
 /note="Vector: pYX-ASC"
 1..3028
 /gene="Tyki"
 /note="synonym: TDKI"
 /db_xref="locusID:22169"
 /db_xref="MGI:99830"
 41..1294
 /gene="Tyki"
 /codon_start=1
 /product="Tyki protein"
 /protein_id="AAH57565.1"
 /db_xref="GI:34785825"
 /db_xref="LocusID:22169"
 /db_xref="MGI:99830"
 /translation="MARPRFTVLPDCLSHFLVIGDATHDRDLARLAELLPPGRSVA
 LCVPLAFEGCGPRVQARVHRLLOLRRLRGLCOLSLGYCGPDGAGEAQHGL
 LRDPCHDPTRRDLQLLGSQAEARPLAEFOADSGLLWRLWELQGRQVQVDC
 CVLPACRPHLPLLDLLNSAVFQDRDAARVALECSFIPKRAVLDLVDQCPKEVQ
 KGRFQVTAIEGLDAGTKITLITVSSELSKAVLIQSPPCISQWRKIFDDEPTIIRAF
 YSLGNYLVAEIEAKESNFVPIVDRIYWHSTAYIAIEVSGGLQYLPFAHPVYQWPG
 DLLKPDVLLVLTWNSEVRVLRQGRQEKTEEALEANNVFRKQVEMTYQRWENPSC
 HLVDASPSRETVLQKVELIQSSGR"

gene

CDS

ORIGIN

Query Match 34.0%; Score 873; DB 10; Length 3028;
 Best Local Similarity 77.2%; Pred. No. 6.5e-132;
 Matches 1105; Conservative 0; Mismatches 310; Indels 16; Gaps 3;
 70 GGGGCCCTGCTGGGGCGGCGGGGCTGGCTGGGGCCATGGCTCGCGCTCCGCTT 129
 Db 1 CGGAGAGCTGCGGGGCGACACAGGCTGTCATGGGACCATGGCTCGGGCGGCGCTT 60
 130 CGTCTGGAGCTCCCGACTGACCCCTGGGTCACTTGGCCCTAGGCGCGACGCCCGCG 189
 Db 61 CACTGTGGAGTGCAGATGCTCCCTGACTCACTTGTCTCTGG-----GGA 108
 190 CGAGGAGAGCGCCCGACCCCGCTGGCGGCGCTGTGGGGCCCGCGAGCGAGCTA 249
 Db 109 TGCAACAGACACAGGATGACGCTTGGCAGAGCTGCTGGGCCCGCGAGGCGCAGCTA 168
 250 CTGCTGTGGTGGCTGACCCCGAGCGCGGCTGGGGCGGCGCGGCTCGGGCGGCGCG 309
 Db 169 CGGCTGTGTGGCTGGCTCAGGCGAAGGCTCGGGCGGCGGCGGCTGAGGCGGCGCG 228
 310 GTGACACAGCGCTGTGACACAGCTGCGCGCGGCGGCGGCTTCCAGCGGTGCCAGCTGCT 369
 Db 229 GGTGACCATCGCTGTGCGAGAGCTGCGCGGGTCTTTACAGAGATGCCAATGAG 288
 370 CAGGCTGCTGTGCTACTGCGGGCGGCGAGCGCGGCGGCGGCGAGCAAGAGCTTCTGCT 429
 Db 289 CAAGCTTCTGGGCTATGCTCGGGCGATCAGCGCGGCGGCGGCGGCTGCTCTCTGCT 348
 430 GCGGACCCCTGATGACCTGACACCGCGAGCGCTGCTCGAGCTGTGGGCGGCGCTG 489
 Db 349 GCGGACCCCTTGGACACCGCGAGCTGCGGCGGCGGCTTGTCTGAGCTTCTGGCTCTG 408
 490 CCAGGAGGACACCGCGCGGCTTGGGCGAGTTCCAGGCGGACCGCGCGGCGGCGAGCTGTG 549
 Db 409 CCAGGAGGCGGCGGCGGCGGCTTGGGCGAGTTCCAGGCGGACTCCAGGCTTGTCTGTG 468

Search completed: October 22, 2004, 18:08:31
 Job time : 10852 secs

This Page Blank (uspto)